Stochastic Calculus Notes[1]

Sergio Ocampo Díaz

University of Minnesota

Contents

[1]These notes are intended to summarize the main concepts, definitions and results covered in the course "dynamic games, contracts and markets" for the Summer School at Pontificia Universidad Javeriana. The material is not my own, these notes only include selected sections of books or articles relevant to the course. Please let me know of any errors that persist in the document. E-mail: ocamp020@umn.edu.

Part I
Random Variables and Probability

Consider an experiment that can have several (but finite) outcomes. For example trowing a dice can turn out in getting any number from 1 to 6, or asking someone out can generate an affirmative response, a negative one or perhaps a maybe, or no response at all. A probability function is a function that assigns a value to each possible outcome while satisfying certain rules.

Its clear that since the outcomes here are finite, outcomes form a set $S = \{s_1, \ldots, s_n\}$ a probability is then a list $(\pi_1, \ldots, \pi_n)$ such that $\Pr(s_i) = \pi_i$:

i. $\pi_i \geq 0$ for all i.

ii. $\sum \pi_i = 1$.

Note that it is natural to define other outcomes that are formed by unions of the former ones, like getting an even number when trowing the dice (the union of getting a two a four and a six) or getting a positive answer or a maybe when asking someone out. It is clear that the probability of these new outcomes is defined by the sum of probabilities of the original outcomes used to define them.

Formally we could say that for any set $A \subseteq S$ we define $I_A = \{i | s_i \in A\}$ and then a function $\mu : 2^S \to [0,1]$ as:

$$\mu(A) = \Pr(A) = \sum_{i \in I_A} \pi_i$$

Furthermore we can define the expected value of a real valued function $f : S \to \mathbb{R}$ as $E[f] = \sum \mu(\{s_i\}) f(s_i)$.

This same discussion can be carried out if the possible outcomes are countably infinite, but it is difficult to generalize it otherwise. The objective now is to study which properties does this kind of function satisfy and how it is generalized to deal with cases where outcomes are arbitrary. The key for this is to realize that a probability is a function that maps sets into the interval $[0,1]$, hence the study of functions that map sets into non-negative numbers will provide the necessary theory, these functions are called measures, for obvious reasons.

The following sections draw on the short exposition of measure theory contained in Chapter 7 of Stokey et al. (1989) and complements it with portions of Kolmogorov and Fomin (2012) (chapters 7 to 10). Both these references are introductory although they present all the relevant results. All the material is also covered in a more advanced manner in Kolmogorov and Fomin (1999).

The aim of the course is not to dwell in the mathematical details of the theory but rather present the most useful results for applications in economic theory, because of this many of the proofs will be omitted only including those that are either instructive of the way the theory is developed. Kolmogorov and Fomin (2012) is a good source for detailed (and easy to understand) proofs.

Markov processes are defined following Stokey et al. (1989), chapter 8, and we finish with the definition of the most common stochastic processes used in the rest of the course.

1 Measure

1.1 Measurable spaces (σ-algebras)

Before we define a measure recall that a measure has for domain a collection of sets. For a measure to have some desirable properties this collection of sets cannot be left unrestricted. It turns out that the appropriate family of sets to be consider is that of σ-algebra.

Definition 1.1. **(σ-algebra)** Let S be a set and $\mathcal{A} \subseteq 2^S$ a family of its subsets. $\mathcal{A}$ is a σ-algebra if and only if:

i. $\emptyset, S \in \mathcal{A}$.

ii. $A \in \mathcal{A}$ implies $A^c = S \backslash A \in \mathcal{A}$. We say that $\mathcal{A}$ is closed under complement.

iii. $A_n \in \mathcal{A}$ for $n = 1, \ldots$ implies $\cup A_n \in \mathcal{A}$. We say that $\mathcal{A}$ is closed under countable union.

    (a) Since $\cap A_n = (\cup A_n^c)^c$ we have that $\mathcal{A}$ is closed under countable intersection.

If $\mathcal{A}$ is only closed under finite union (or intersection) then $\mathcal{A}$ is an algebra.

A σ-algebra imposes certain consistency to the family of sets under consideration. The way to interpret it is that only subsets of the σ-algebra can be known, hence measured. Because of property (i) it is possible to know when none or all of the outcomes occurred. Also if there is an outcome that occurred it must be possible to determine if it didn't. Finally if it is possible to determine that some outcomes occurred individually it can also be determined if at least one or all of them were realized.

It is instructive to consider two simple examples of σ-algebras that arise from throwing a 4 sided dice, then $S = \{1, 2, 3, 4\}$. One (trivial) σ-algebra is:

$$\mathcal{A} = \{\emptyset, S\}$$

Another one is the σ-algebra generated by the collection $\{\{1\}, \{2\}, \{3\}, \{4\}\}$, then:

$$\mathcal{A} = \left\{ \begin{array}{c} \{1\}, \{2\}, \{3\}, \{4\}, \{2,3,4\}, \{1,3,4\}, \{1,2,4\}, \{1,2,3\}, \\ \{1,2\}, \{2,3\}, \{3,4\}, \{1,3\}, \{1,4\}, \{2,4\}, \emptyset, S \end{array} \right\}$$

In this case $\mathcal{A} = 2^S$, but this is not necessarily true, imagine that one can only determine if an even number was thrown, then the outcomes are $\{\{1,3\}, \{2,4\}\}$, the σ-algebra is:

$$\mathcal{A} = \{\{1,3\}, \{2,4\}, \emptyset, S\}$$

When S has uncountably many elements this process cannot be exemplified as easily but one can always define the σ-algebra generated by a subset $\mathcal{A} \subseteq 2^S$ as the intersection of all σ-algebras that contain $\mathcal{A}$. Clearly the arbitrary intersection of σ-algebras is again a σ-algebra.

Now that we have defined a σ-algebra its possible to say what a measurable set and a measurable space are:

Definition 1.2. (Measurable Space) A pair $(S, \mathcal{A})$ where S is any set and $\mathcal{A}$ is a σ-algebra is called a measurable space. A set $A \in \mathcal{A}$ is called $\mathcal{A}$-measurable.

Note that we say that $A \subseteq S$ is measurable with respect to a σ-algebra $\mathcal{A}$ if its elements are identifiable, that is, if the outcomes represented in A can be told apart from other outcomes given the information in $\mathcal{A}$. For example the set $A = \{4\}$ is not measurable in the last example above, since its impossible to know if a 4 was the outcome of the throw.

A σ-algebra of special importance is the Borel σ-algebra.

Definition 1.3. (Borel σ-algebra) Let $S = \mathbb{R}$ and $\mathcal{A}$ be the set of open and half open intervals. The Borel algebra, noted by $\mathcal{B}$, is the σ-algebra generated by $\mathcal{A}$. A set $B \in \mathcal{B}$ is called a Borel set.

Note that the Borel algebra could have been defined equivalently with the closed and half closed intervals (use complement). In general one can define the Borel algebra for any metric space (S, ρ) as the smallest σ-algebra containing all the open balls. In the case of the Euclidean spaces it can also be generated with open rectangles.

What follows is to define the measure of a measurable set.

1.2 Measures

1.2.1 Measures in σ-algebras

Given a measurable space $(S, \mathcal{A})$ a measure is nothing but a function $\mu : \mathcal{A} \to \overline{\mathbb{R}}$ with certain restrictions that guarantee its consistency:

Definition 1.4. (Measure) Let $(S, \mathcal{A})$ be a measurable space. A measure is an extended real-valued function $\mu : \mathcal{A} \to \overline{\mathbb{R}}$ such that:

i. $\mu(\emptyset) = 0$

ii. $\mu(A) \geq 0$ for all $A \in \mathcal{A}$.

iii. μ is countably additive. If $\{A_n\}_{n=1}^{\infty}$ is a countable, disjoint sequence in $\mathcal{A}$, then:

$$\mu(\cup A_n) = \sum \mu(A_n)$$

If furthermore $\mu(S) < \infty$ then μ is said to be a finite measure, and if $\mu(S) = 1$ then μ is said to be a probability measure.

Definition 1.5. (Measure Space) A triple $(S, \mathcal{A}, \mu)$ where S is a set, $\mathcal{A}$ is a σ-algebra of its subsets and μ is a measure on $\mathcal{A}$ is called a measure space. The triple is called a probability space if μ is a probability measure.

An important concept is that of almost everywhere and almost surely. These are qualifiers for a given proposition that can be evaluated in sets of $\mathcal{A}$.

Definition 1.6. (Almost Everywhere and Almost Surely) Let $(S, \mathcal{A}, \mu)$ be a measure space. A proposition is said to hold almost everywhere (a.e.) or almost surely (a.s.) if there exists a set $A \in \mathcal{A}$ such that $\mu(A) = 0$ and the proposition holds in A^c.

An example of the use of a.e. or a.s. is when treating functions that are similar to each other. One can say that two functions are equivalent a.e. or that a function is continuous a.e. Then the functions f and g satisfy $f(x) = g(x)$ and $A = \{x | f(x) \neq f(y)\}$ satisfies $\mu(A) = 0$. In measure theory the behavior of functions a.e. is all that matters, then we can treat functions that have anomalies as long as those anomalies occur only in sets of measure zero.

There are some properties of a measure that are useful to keep in mind, a crucial one is used for Bayes law and the definition of conditional probability.

Proposition 1.1. *Let $(S, \mathcal{A}, \mu)$ be a measure space and $B \in \mathcal{A}$ a set. Define $\lambda : \mathcal{A} \to \overline{\mathbb{R}}$ as $\lambda(A) = \mu(A \cap B)$. Then λ is a measure on $(S, \mathcal{A})$. If in addition $\mu(B) < \infty$ then $\tilde{\lambda}$ defined as $\tilde{\lambda}(A) = {}^{\mu(A\cap B)}/_{\mu(B)}$ is a probability measure on $(S, \mathcal{A})$.*

Proof. First note that if $A, B \in \mathcal{A}$ then $A \cap B \in \mathcal{A}$, this follows from a σ-algebra being closed under countable intersection, by letting $A_1 = A$ and $A_n = B$ for $n \geq 2$ the result obtains. It is left to check the three properties of a measure:

i. $\lambda(\emptyset) = \mu(\emptyset \cap B) = \mu(\emptyset) = 0$.

ii. $\lambda(A) = \mu(A \cap B) \geq 0$.

iii. Let $\{A_n\}_{n=1}^{\infty}$ be a countable, disjoint sequence in $\mathcal{A}$, then note that the sequence $\{A_n \cap B\}_{n=1}^{\infty}$ is also disjoint and that:

$$\lambda(\cup A_n) = \mu((\cup A_n) \cap B) = \mu(\cup (A_n \cap B)) = \sum \mu(A_n \cap B) = \sum \lambda(A_n)$$

iv. If $\mu(B) < \infty$ then all the previous results hold for $\tilde{\lambda}$ by dividing everything by $\mu(B)$. Furthermore $\tilde{\lambda}(S) = \frac{\mu(S \cap B)}{\mu(B)} = \frac{\mu(B)}{\mu(B)} = 1$.

□

Another useful property is given by the following proposition, it reflects the intuitive property of measures being 'increasing':

Proposition 1.2. *Let $(S, \mathcal{A}, \mu)$ be a measure space and $A, B \in \mathcal{A}$ sets. If $A \subseteq B$ then $\mu(A) \leq \mu(B)$, if in addition μ is finite then $\mu(B \backslash A) = \mu(B) - \mu(A)$.*

Proof. Since $A \subseteq B$ there exits $C = B \backslash A = B \cap A^c$ such that $A \cup C = B$ and $A \cap C = \emptyset$. Then:

$$\mu(A) + \mu(C) = \mu(B)$$

since $\mu(C) \geq 0$ if follows that $\mu(A) \leq \mu(B)$. If μ is finite then all elements above are well defined and: $\mu(B \backslash A) = \mu(B) - \mu(A)$. □

The following property is widely used to establish properties of limits of functions, and of the Lebesgue integral:

Proposition 1.3. *Let $(S, \mathcal{A}, \mu)$ be a measure space:*

i. If $\{A_n\}$ is an increasing sequence in $\mathcal{A}$, that is, if $A_n \subseteq A_{n+1}$ for all n, then:

$$\mu(\cup A_n) = \lim \mu(A_n)$$

ii. If $\{B_n\}$ is an decreasing sequence in $\mathcal{A}$, that is, if $B_n \supseteq B_{n+1}$ for all n, then:

$$\mu(\cap B_n) = \lim \mu(B_n)$$

Proof. Stokey et al. (1989, Sec. 7.2). Satisfying these two properties makes a measure continuous. □

1.2.2 Measures in algebras and extensions [Optional]

So far we have defined a measure on an σ-algebra, but a σ-algebra is usually a large collection of sets and defining a function on such a set while preserving the consistency required for a measure is not an easy task. An alternative is given by defining measures on algebras, which are smaller and less complicated collections of sets. It can be shown that these measures preserve all the desirable properties of the more complicated spaces, and also allow for an extension to σ-algebras, once the measure is properly constructed.

We start by defining a measure on an algebra.

Definition 1.7. (Measure) Let $(S, \mathcal{A})$ be a measurable space. A measure is an extended real-valued function $\mu : \mathcal{A} \to \overline{\mathbb{R}}$ such that:

i. $\mu(\emptyset) = 0$

ii. $\mu(A) \geq 0$ for all $A \in \mathcal{A}$.

iii. If $\{A_n\}_{n=1}^{\infty}$ is a countable, disjoint sequence in $\mathcal{A}$, and $\cup A_n \in \mathcal{A}$, then:

$$\mu(\cup A_n) = \sum \mu(A_n)$$

If furthermore $\mu(S) < \infty$ then μ is said to be a finite measure, and if $\mu(S) = 1$ then μ is said to be a probability measure.

Note that condition (iii) also includes finite union of disjoint sets as a special case.

Definition 1.8. (σ-finite measure) Let S be a set, $\mathcal{A}$ an algebra of its subsets and μ a measure defined on $\mathcal{A}$. If there is a countable sequence of sets in $\mathcal{A}$, $\{A_n\}$, such that $\mu(A_n) \leq \infty$ and $S = \cup A_n$ then μ is σ-finite

It is now possible to extend the notion of this measure to a σ-algebra.

Theorem 1.1. (Caratheodory extension theorem) Let S be a set, $\mathcal{A}$ an algebra of its subsets and μ a measure defined on $\mathcal{A}$. Let $\mathcal{A}^{\star}$ be the smallest σ-algebra containing $\mathcal{A}$. There exists a measure $\mu^{\star}$ on $\mathcal{A}^{\star}$ such that $\mu^{\star}(A) = \mu(A)$ for all $A \in \mathcal{A}$.

The problem of uniqueness is also solved.

Theorem 1.2. (Hahn extension theorem) Let S be a set, $\mathcal{A}$ an algebra of its subsets, μ a measure defined on $\mathcal{A}$ and $\mathcal{A}^{\star}$ the minimal σ-algebra of $\mathcal{A}$. If μ is σ-finite then the extension $\mu^{\star}$ is unique.

To see how these theorems and the extension of a measure are used consider defining a measure on the Borel σ-algebra. It seems logical to define the measure of an interval $A = (a, b)$ as $\mu(A) = b - a$ if $b \geq a$ and $\mu(A) = 0$ otherwise (since the interval would be empty). Yet the Borel σ-algebra contains sets beyond simple intervals, and the countable union of intervals can give rise to weird sets. An answer to this problem is given by defining a measure on the Borel algebra, formed by all types of intervals and their finite unions. Defining a measure on this set seems straightforward:

i. $\mu(\emptyset) = 0$

ii. $\mu((a,b)) = \mu([a,b]) = \mu((a,b]) = \mu([a,b)) = b - a$

iii. $\mu((-\infty,\infty)) = \mu((-\infty,b]) = \mu([a,\infty)) = \infty$

iv. $\mu(\cup(a_n,b_n)) = \sum(b_n - a_n)$ if the intervals are disjoint.

The function μ can be verified to be a measure on the Borel algebra, and hence an extension to the Borel σ-algebra exists. If we restrict our attention to $S = [a,b]$ and the intervals contained in it we can define a σ-finite measure, obtaining uniqueness of the extension. This is how we can deal with complicated environments.

Once the measure is extended to the σ-algebra all the results obtained above apply.

1.2.3 Completion of a measure [Optional]

One small detail is left to be checked. Sometimes there is a set $B \subseteq S$ such that $B \subseteq A \in \mathcal{A}$ and $\mu(A) = 0$, but if $B \notin \mathcal{A}$ then its measure is undefined, while it should be clearly zero. The completion of a σ-algebra to include these type of 'harmless' sets is what follows. Note that as before including sets or behaviors of measure zero is of no consequence.

Definition 1.9. (Completion of a σ-algebra) Let $(S,\mathcal{A},\mu)$ be a measure space. Define a collection $\mathcal{C}$ as:

$$\mathcal{C} = \{C \subset S | \exists_{A\in\mathcal{A}} \mu(A) = 0 \quad \wedge \quad C \subset A\}$$

The completion of σ-algebra $\mathcal{A}$ is:

$$\mathcal{A}' = \left\{B' \subseteq S \,|\, B' = (A \cup C_1) \setminus C_2 \qquad A \in \mathcal{A} \quad \wedge \quad C_1, C_2 \in \mathcal{C}\right\}$$

Note that by letting $C_1 = C_2 = \emptyset$ we get $\mathcal{A} \subseteq \mathcal{A}'$, $\mathcal{A}'$ includes all sets in 2^S that differ from a set in $\mathcal{A}$ by a set of measure 0.

Definition 1.10. (Completion of a measure) Let $(S,\mathcal{A},\mu)$ be a measure space and $\mathcal{A}'$ the completion of $\mathcal{A}$. $\mu(B') = \mu(B)$ for any $B' \in \mathcal{A}'$ that differs from $B \in \mathcal{A}$ by a set of measure 0.

The Caratheodory and Hahn extension theorems also apply for completions.

2 Measurable functions

A measurable function is a type of function for which it is possible to know (to measure) the conditions (the set) that originates certain outcomes. One can think of a function as mapping certain events in a given measure space to outcomes in another measure space. A function is measurable if the sets that induce a given outcome are measurable. Formally:

Definition 2.1. (Measurable function) Let $(S, \mathcal{A}, \mu)$ and $\left(S^{'}, \mathcal{A}^{'}, \mu^{'}\right)$ be measure spaces and $f : S \to S^{'}$ a function. f is measurable if and only if $f^{-1}\left(A^{'}\right) \in \mathcal{A}$ for all $A^{'} \in \mathcal{A}^{'}$.

A special case of notable importance is that of $\left(S^{'}, \mathcal{A}^{'}, \mu^{'}\right) = (\mathbb{R}, \mathcal{B}, \lambda)$, where λ is the Lebesgue measure on the plane. This are real valued functions. In this case the $\mathcal{B}$-measurable sets in $\mathbb{R}$ can be characterized in the following way:

Theorem 2.1. *Let $(S, \mathcal{A}, \mu)$ be a measure space and $f : S \to \mathbb{R}$. f is μ-measurable if and only if $f^{-1}((-\infty, c)) = \{x \in S | f(x) < c\} \in \mathcal{A}$ for all $c \in \mathbb{R}$.*

Proof. This theorem is stated as the definition of a real valued function f being μ-measurable in Stokey et al. (1989), but a formal proof is presented in Kolmogorov and Fomin (2012, Sec. 28, Thm. 1). It can also be stated with any of the inequalities $\geq, \leq, >, <$. $\square$

Also when the measure space in question is a probability space one can characterize formally what a random variable is.

Definition 2.2. (Random variable) Let $(S, \mathcal{A}, P)$ be a probability space and $f : S \to \mathbb{R}$ a real valued function. f is a random variable if and only if f is measurable, that is, if and only if $f^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}$, where $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{R}$. We further establish the same notation:

i. An outcome is an element $s \in S$.

ii. An event is a measurable subset of S: $A \in \mathcal{A}$.

iii. The real number $f(s)$ is a realization of the random variable.

iv. The probability measure for f is then: $\mu(B) = P\left(f^{-1}(B)\right) = P(\{s \in S | f(s) \in B\})$, for $B \in \mathcal{B}$.

v. The distribution function for f is: $G(b) = \mu((-\infty, b])$, for $b \in \mathbb{R}$.

Generally it is very hard to find a function that is not measurable. The details of the example will depend on the spaces considered. For example if $f : \mathbb{R} \to \mathbb{R}$ and $\mathcal{A}$ is the set of all open (or closed) sets in $\mathbb{R}$ the definition of measurability is equivalent to that of continuity (the pre-image of an open set has to be open) and then all functions that are not continuous are not measurable. It is clear that more complete σ-algebras make more difficult to generate counterexamples. The following three results show how difficult it is to generate them:

Proposition 2.1. *Let $f : \mathbb{R} \to \mathbb{R}$.*

i. If f is continuous then f is measurable with respect to the Borel sets.

ii. If f is monotone then f is measurable with respect to the Borel sets.

Proof. Each case is proven:

i. Let f be continuous. Consider the set $f^{-1}((-\infty, c))$ for any $c \in \mathbb{R}$, note that $(-\infty, c)$ is open, since f is continuous then its pre-image is open, then it is a Borel set. Then its measurable.

ii. Let f be monotone increasing. Consider the set $f^{-1}((-\infty, c))$ for arbitrary $c \in \mathbb{R}$. Note that $f^{-1}((-\infty, c)) = (-\infty, a)$ or $f^{-1}((-\infty, c)) = (-\infty, a]$ or $f^{-1}((-\infty, c)) = (-\infty, \infty)$ or $f^{-1}((-\infty, c)) = \emptyset$ for some $a \in \mathbb{R}$. Monotonicity ensures that if $a \in f^{-1}((-\infty, c))$ and $b \leq a$ then $b \in f^{-1}((-\infty, c))$. Suppose its not, then there exists numbers $b \leq a$ such that $f(b) > c \geq f(a)$, contradicting monotonicity.

   Note that all these sets are in $\mathcal{B}$, then f is $\mathcal{B}$-measurable.

$\square$

Corollary 2.1. *The composition of measurable functions is measurable. In particular the composition of a continuous function with a measurable function is measurable.*

Proposition 2.2. *Let $S = \{s_1, s_2, \ldots\}$ be a countable set (potentially infinite) and $\mathcal{A} = 2^S$ a σ-algebra on S. Then all functions $f : S \to \mathbb{R}$ are measurable.*

Proof. The proof is immediate since the pre-image of a Borel set is a subset of S, then it belongs to $\mathcal{A} = 2^S$. $\square$

In a more general way one can establish the measurability of a function by relating to a class of well behave 'simple' functions. The base for this class is the indicator function.

Definition 2.3. (Indicator Function) Let $(S, \mathcal{A})$ be a measurable space. An indicator function $\chi_A : S \to \mathbb{R}$ is:

$$\chi_A(s) = \begin{cases} 1 & \text{if } s \in A \\ 0 & \text{if } s \notin A \end{cases}$$

Clearly χ_A is measurable if and only if $A \in \mathcal{A}$.

Definition 2.4. (Simple Function) Let $(S, \mathcal{A})$ be a measurable space. A simple function is a function that takes at most countably many values. When the function takes finitely many values it can be expressed as:

$$\phi(s) = \sum_{i=1}^{n} \alpha_i \chi_{A_i}(s)$$

where $\{A_i\}$ is a sequence of subsets of S and $\alpha_i \in \mathbb{R}$.

Characterizing the measurability of simple functions is slightly more complicated.

Proposition 2.3. *A simple function taking values $\{y_1, y_2, \ldots\}$ is measurable if and only if the sets $A_i = \{s \in S | \phi(s) = y_n\}$ are measurable.*

Proof. Both directions are proven.

i. Let ϕ be measurable, and note that $\{y_n\} \in \mathcal{B}$, then its pre-image is measurable wrt $\mathcal{A}$.

ii. Let the sets be measurable, that is$A_i \in \mathcal{A}$, and consider $B \in \mathcal{B}$ a Borel set. Then:

$$\phi^{-1}(B) = \{s \in S | \phi(s) = y_i \in B\} = \bigcup_{y_i \in B} A_i$$

Since each $A_i \in \mathcal{A}_i$ and the union is taken over no more than countably many sets we have $\bigcup_{y_i \in B} A_i \in \mathcal{A}$ by definition of a σ-algebra. This proves measurability of $\phi^{-1}(B)$.

□

In what follows all simple functions will be considered measurable. The importance of simple functions is given by the applications of the following proposition.

Proposition 2.4. *Let $(S, \mathcal{A})$ be a measurable space and let $\{f_n\}$ be a sequence of measurable functions converging pointwise to f, that is $\lim f_n(s) = f(s)$ for all s. Then f is also measurable.*

Proof. The proof can be found in Stokey et al. (1989, Sec. 7.3) or in Kolmogorov and Fomin (2012, Sec. 28.1). □

Corollary 2.2. *If f is non-negative one can choose the sequence $\{f_n\}$ to be strictly increasing.*

Corollary 2.3. *If f is bounded one can choose the sequence $\{f_n\}$ to converge uniformly.*

The main application is the following result that gives a characterization of measurable functions in terms of simple functions:

Proposition 2.5. *A function f is measurable if and only if it an be represented as the limit of a uniformly converging sequence of measurable simple functions.*

Proof. The first direction is immediate from the previous proposition. If f is the limit of measurable functions then f is also measurable.

Let f be measurable. It is left to construct a converging sequence of simple functions that converges to f. wlog let $f(s) \geq 0$ for all s, then by the Archimedean principle there exists a non-negative integer m such that

$$\frac{m}{n} \leq f(s) < \frac{m+1}{n}$$

Let $f_n(s) = {}^m\!/_n$, since n is fixed and $m \in \mathbb{N} \cup \{0\}$ it follows that f_n can take at most countably many values, hence it is simple. f_n is also measurable since:

$$f_n^{-1}((-\infty, c)) = \{s \in S | f_n(s) \leq c\} = \left\{s \in S | f_n(s) \leq \frac{m^\star}{n}\right\} = \left\{s \in S | f_n(s) < \frac{m^\star + 1}{n}\right\}$$

For $m^\star$ chosen by the Archimedean principle. Note that the last set is $f^{-1}\left(\left(-\infty, \frac{m^\star+1}{n}\right)\right)$ which is measurable by assumption. Then f_n is measurable for all n.

Finally note that $f_n \to f$ uniformly since:

$$\left|f_n\left(s\right) - f\left(s\right)\right| \leq \left|\frac{m}{n} - \frac{m+1}{n}\right| = \frac{1}{n}$$

□

Other results will follow and are left stated without proof:

Proposition 2.6. *Let f, g be measurable functions and $\alpha \in \mathbb{R}$ then:*

i. *$f + g$ is measurable.*

ii. *αf is measurable.*

iii. *fg is measurable.*

iv. *$1/f$ is measurable provided that $f\left(s\right) \neq 0$.*

Finally continuity of functions is used to strengthen the intuition around measurability.

Proposition 2.7. *Let f, g be equivalent function defined on an interval E, that is they are equal a.e. If f and g are continuous then they coincide.*

Proof. Suppose not, then there exists $x \in E$ such that $f\left(x\right) \neq g\left(x\right)$. Let $\epsilon = \left|f\left(x\right) - g\left(x\right)\right|$, since f and g are continuous there exists δ such that for $x^{'} \in B_{\delta}\left(x\right)$ it holds that $\left|f\left(x\right) - f\left(x^{'}\right)\right| < \frac{\epsilon}{2}$ and $\left|g\left(x\right) - g\left(x^{'}\right)\right| < \frac{\epsilon}{2}$. Then for all $x^{'} \in B_{\delta}\left(x\right)$ it holds that $f\left(x^{'}\right) \neq g\left(x^{'}\right)$, but $B_{\delta}\left(x\right)$ has strictly positive measure, contradicting f and g being equivalent. □

Proposition 2.8. *A function f equivalent to a measurable function g is measurable.*

Proof. Since the functions are equivalent the sets $\{x|f\left(x\right) \leq c\}$ and $\{x|g\left(x\right) \leq c\}$ can differ in at most by a set of measure zero. Then if the second set is measurable so is the first one (taking into account the completion of the σ-algebra). This proves measurability. □

Corollary 2.4. *A function f equivalent to a continuous function is measurable.*

Proof. Immediate from continuous functions being measurable. □

This implies that if a function is continuos a.e. then it is measurable, again the behavior of functions in sets of measure zero carries no consequence. It turns out that this corollary can be strengthened. The result is powerful and is stated without a proof:

Theorem 2.2. ***(Luzin)*** *Let $f : [a, b] \to \mathbb{R}$ be a function. f is continuous if and only if for all $\epsilon > 0$ there exists a continuous function g such that $\mu\left\{x \in [a, b] \mid f\left(x\right) \neq g\left(x\right)\right\} < \epsilon$.*

This theorem shows that for the case of functions of real variable and real value measurability is equivalent to continuity, except on a set of arbitrarily small size. In other words a measurable function can be made continuous by altering its values on a set of arbitrarily small measure.

3 The Lebesgue integral

The Lebesgue integral is in at least two important ways a generalization of the Riemann integral and it serves a crucial purpose of defining what it means to take the expected value of a function with respect to a probability distribution. The first sense in which the Riemann integral is generalized is that the Lebesgue integral is defined over measurable functions, a space that is much richer than that of Riemann integrable functions, the second sense is much more crucial: the Lebesgue integral is defined for functions with domain in arbitrary sets, thus allowing to handle a more abstract and general class of functions.

Intuitively the Lebesgue integral is constructed in a similar way than the Riemann integral. To construct the latter one takes successively finer grids of the domain and evaluate the function at certain points, constructing step functions, one above the function and one below, then two sums are constructed and the value of the integral is defined as the (common) value of the limit of those sums as the length of the grid's spaces goes to zero.

The Lebesgue integral of a function $f : S \to \mathbb{R}_+$ is constructed by taking grids over the range of the function $\{y_i\}_{i=1}^n$ such that $0 = y_1 \leq \ldots \leq y_n$. Then one can define the sets $A_i = \{s \in S | y_i \leq f(s) < y_{i+1}\}$ and using the measure over S define $\lambda(A_i)$ and the sum $\sum y_i \lambda(A_i)$. The Lebesgue integral is then the limit of this sum as the values y_i are closer together.

The introduction before of simple functions makes sense when defining the Lebesgue integral. Its definition seems intuitive for this class of functions and Proposition 2.5 creates a bridge between them and the more general class of measurable functions, thus allowing to extend the Lebesgue integral to this broader family.

In what follows we restrict attention to non-negative, real valued functions.

Definition 3.1. (Lebesgue integral for simple functions) Let $(S, \mathcal{A}, \mu)$ be a measure space and $f : S \to \mathbb{R}_+$ a simple, μ-measurable function that takes no more than countably many values $\{y_1, y_2, \ldots\}$. The Lebesgue integral over the set $A \subseteq S$ is defined as:

$$\int_A f(s)\, d\mu = \sum_n y_n \mu(A_n) \tag{3.1}$$

where the sets A_n are defined as:

$$A_n = \{s \in A | f(s) = y_n\}$$

Note that these sets can be empty if there is no element of s in A for which f takes a given value. The Lebesgue integral is defined as long as the series in (3.1) is absolutely convergent. Note that if f takes finitely many values and μ is finite (or a probability measure) this condition is satisfied.

An example is given by the constant function, $f(s) = 1$ for all $s \in S$, then:

$$\int_A f(s)\, d\mu = \int_A d\mu = \mu(A)$$

It can be shown that the lebesgue integral satisfies some natural properties:

Proposition 3.1. *Let f and g be non-negative, measurable, simple and integrable functions on $(S, \mathcal{A}, \mu)$, a measure space, and $c \geq 0$ a constant. Then:*

i. $\int_A (f+g)(s)\, d\mu = \int_A f(s)\, d\mu + \int_A g(s)\, d\mu$

ii. $\int_A (cf)(s)\, d\mu = c \int_A f(s)\, d\mu$

iii. If f is bounded $|f(s)| \leq M$ a.e. then f is integrable and $\left|\int_A f(s)\, d\mu\right| \leq M\mu(A)$.

Proof. Kolmogorov and Fomin (2012, Sec. 29.1). □

Definition 3.2. (Lebesgue integral - Nonnegative functions) Let $(S, \mathcal{A}, \mu)$ be a measure space. A measurable function $f : S \to \mathbb{R}$ is said to be integrable on a set A if there exists a sequence $\{f_n\}$ of integrable simple functions converging uniformly to f on A. The Lebesgue integral is defined as:

$$\int_A f(s)\, d\mu = \lim \int_A f_n(s)\, d\mu \tag{3.2}$$

Note that this definition precludes the integral from being infinite, as shown in Kolmogorov and Fomin (2012, Sec. 29.1), the limit above exists provided that the functions f_n are integrable (recall that it was asked of the sum in (3.1) to be finite), moreover it is independent of the choice of sequence approximating f, this sequence can be furthermore be chosen to be strictly increasing (Stokey et al., 1989). Yet, the concept of the Lebesgue integral can be easily generalized to allow for infinite values, the definition in Stokey et al. (1989) allows for this.

What follows is a list of properties of the Lebesgue integral which should be familiar if there is any knowledge of the behavior of Riemann integrals. They are not of particular interest in this course.

Proposition 3.2. *Properties of the Lebesgue integral for non-negative measurable functions:*

i. $\int_A (f+g)(s)\, d\mu = \int_A f(s)\, d\mu + \int_A g(s)\, d\mu$

ii. $\int_A (cf)(s)\, d\mu = c \int_A f(s)\, d\mu$

iii. If g is measurable and integrable and f is bounded by g: $|f(s)| \leq g(s)$ a.e., then f is integrable and $\left|\int_A f(s)\, d\mu\right| \leq \int_A g(s)\, d\mu$.

  (a) If f is bounded and measurable then it is integrable.

iv. If $f \leq g$ a.e. then $\int f(s)\, d\mu \leq \int g(s)\, d\mu$.

v. If $A \subseteq B$ with $A, B \in \mathcal{A}$ then $\int_A f(s)\, d\mu \leq \int_B f(s)\, d\mu$

vi. Let $A = \cup A_n$ where $\{A_n\}$ is a finite or countable sequence of disjoint sets. If f is integrable on A then f is integrable on A_n for all n and:

$$\int_A f(s)\, d\mu = \sum_n \int_{A_n} f(s)\, d\mu$$

when the series on the right is absolutely convergent.

Finally it is noted that a non-negative integrable function induces a measure on a space, the following proposition makes this clear.

Proposition 3.3. *Let f be a non-negative, integrable function, then $\lambda : \mathcal{A} \to \mathbb{R}$ defined as:*

$$\lambda\left(A\right) = \int_A f\left(s\right) d\mu$$

is a measure on $(S, \mathcal{A})$.

Definition 3.3. (Lebesgue integral) Let $(S, \mathcal{A}, \mu)$ be a measure space. A measurable function $f : S \to \mathbb{R}$ is said to be integrable if the following two integrals are finite:

$$\int f^{+}\left(s\right) d\mu \qquad \int f^{-}\left(s\right) d\mu$$

where:

$$f^{+}\left(s\right) = \begin{cases} f\left(s\right) & \text{if } f\left(s\right) \geq 0 \\ 0 & \text{if } f\left(s\right) < 0 \end{cases} \qquad f^{+}\left(s\right) = \begin{cases} 0 & \text{if } f\left(s\right) \geq 0 \\ -f\left(s\right) & \text{if } f\left(s\right) < 0 \end{cases}$$

The integral of f is defined as:

$$\int f\left(s\right) d\mu = \int f^{+}\left(s\right) d\mu - \int f^{-}\left(s\right) d\mu \tag{3.3}$$

Recall that when $(S, \mathcal{A}, \mu)$ is a probability space the function f is called a random variable, the definitions above are then the definitions of the expected value of a random variable, this expected value exists when f is integrable, we have seen that a sufficient condition for this is to be bounded a.e. and the measure to be finite, this last condition is satisfied immediately by probability measures.

4 The Stieltjes integral [Optional]

The Lebesgue-Stieltjes integral is a type of integral specially useful in probability theory, because of the resemblance between the Stieltjes measures and probability measures. To introduce the concept consider a real valued random variable that takes values on a closed interval $[a, b]$, this is for example the result of coin toss when catalogued as 0 or 1, the underlying probability space is formed by $S = \{H, T\}$, $\mathcal{A} = \{\emptyset, S, \{H\}, \{T\}\}$ and a probability measure on $\mathcal{A}$, a function $\mu : \mathcal{A} \to [0, 1]$ such that $\mu(\{H\}), \mu(\{T\}) \geq 0$, $\mu(S) = \mu(\{H\}) + \mu(\{T\}) = 1$ and $\mu(\emptyset) = 0$. The random variable is then a function $f : S \to \mathbb{R}$ such that $f(H) = 0$ and $f(T) = 1$. It seems natural to ask what is the probability that $f(s) = 1$, it is of course given by $\mu(T)$, in the same way can ask for the probability that $f(s) \leq c$ for any value c, the function that answers that question is called the cumulative distribution function. In this example we have:

$$F(c) = \Pr(f(s) \leq c) = \begin{cases} 0 & \text{if } c < 0 \\ \mu(H) & \text{if } 0 \leq c < 1 \\ 1 & \text{if } 1 \leq c \end{cases}$$

Since the measure μ is non-negative it is clear that F has to be a non-decreasing function, it is also continuous from the left, moreover it is possible to recover μ from knowledge of F:

$$\mu(H) = F(0) \qquad \mu(T) = 1 - F(0)$$

The Stieltjes measure is a general way of looking at this last step. It treats the problem of inducing a measure from a non-decreasing left continuous function. The application to probability theory is apparent since we deal with the CDF of a random variable, and not directly with its probability measure, as we saw before it is this latter object the one that defines the expected value.

Now we turn to define formally the Stieltjes integral. Let $F : [a, b] \to \mathbb{R}$ be a non-decreasing and left-continuous function. Let $\mathcal{A}$ be an algebra of all subintervals of $[\alpha, \beta)$ (including open, closed and half-open intervals). Define a measure on $\mathcal{A}$ by:

$$\begin{aligned} m(\alpha, \beta) &= F(\beta) - F(\alpha + 0) \\ m[\alpha, \beta] &= F(\beta + 0) - F(\alpha) \\ m(\alpha, \beta] &= F(\beta + 0) - F(\alpha + 0) \\ m[\alpha, \beta) &= F(\beta) - F(\alpha) \end{aligned}$$

Now consider the Lebesgue extension of m, call it μ_F and the σ-algebra of all μ_F-measurable, call it $\mathcal{A}_F$. Note that $\mathcal{A}_F$ contains all subintervals of $[\alpha, \beta)$ and hence all the Borel sets of $[\alpha, \beta)$.

Definition 4.1. (Stieltjes measure) The measure μ_F described above is called the (Lebesgue-)Stieltjes measure and F its generating function.

This concept is easily extended to the whole real line. Some examples show the generality of this type of measure:

Example 4.1. Let $F(x) = x$, then the Stieltjes measure is nothing but the Lebesgue measure on the real line, that is, the extension of the concept of length of an interval.

Example 4.2. Let F be a jump function with discontinuity points $\{x_1, x_2, \ldots\}$ and corresponding jumps $\{h_1, h_2, \ldots\}$. The measure is of course:

$$m(\{x_n\}) = h_n \qquad m(\{x_1, x_2, \ldots\}^c) = 0$$

Then every subset of $[\alpha, \beta)$ is μ_F-measurable since their measure depends only on countable points. Any set A has measure given by:

$$\mu_F(A) = \sum_{x_n \in A} h_n$$

This number exists by assumption. A Stieltjes measure generated by a jump function is called a discrete measure. Note that all discrete random variables have CDF that are jump functions.

Example 4.3. Let F be an absolutely continuous non-decreasing function on $[\alpha, \beta)$. Absolutely continuous functions have a finite derivative a.e. let this derivative be $f = F'$. Then the Stieltjes measure μ_F is defined for all Lebesgue measurable sets and:

$$\mu_F(A) = \int_A f(x)\, dx$$

clearly in this case $\mu_F(\{x\}) = 0$ since $\{x\}$ has Lebesgue measure 0.

The result follows from Lebesgue theorem:

Theorem 4.1. ***(Lebesgue)*** *If F is absolutely continuous on $[a, b]$ then the derivative F' is integrable on $[a, b]$ and:*

$$F(\beta) - F(\alpha) = \int_\alpha^\beta F'(x)\, dx$$

Proof. Kolmogorov and Fomin (2012, Sec. 33, Thm. 6). □

Applying this theorem here we get:

$$m(\alpha, \beta) = m[\alpha, \beta] = m(\alpha, \beta] = m[\alpha, \beta) = \int_\alpha^\beta f(x)\, dx$$

Since f is non-negative and integrable wrt all Lebesgue-measurable subsets of $[a, b]$ $\left(\mathcal{B}_{[a,b]}\right)$ we know by proposition (3.3) that

$$\mu_F(A) = \int_A f(x)\, dx$$

is a measure on $\left([a, b], \mathcal{B}_{[a,b]}\right)$ that coincides with m, since the extension is unique we get that μ_F is the Stieltjes measure we are looking for.

This type of measure is called absolutely continuous and is related to continuous random variables.

Now we can define the integral with respect to a Stieltjes measure:

Definition 4.2. (Lebesgue-Stieltjes integral) Let μ_F be Stieltjes measure with generating function F, and let g be a μ_F-measurable function, then the integral is defined as:

$$\int_a^b g(x)\, dF(x) = \int_{[a,b]} g(x)\, d\mu_F$$

If μ_F is discrete with $F(x) = \sum_{x_n \leq x} h(x_n)$, then we have:

$$\int_a^b g(x)\, dF(x) = \sum_n g(x_n)\, h_n$$

If μ_F is absolutely continuous then:

$$\int_a^b g(x)\, dF(x) = \int_a^b g(x)\, f(x)\, dx$$

As hinted above in probability Stieltjes measures arise naturally. Let ξ be a random variable and define $F(x) = \Pr(\xi < x)$, then as noted above F is non-decreasing and continuous from the left, moreover $F(-\infty) = 0$ and $F(\infty) = 1$. The Lebesgue-Stieltjes measure allows us to define the expected value and variance of the random variable as:

$$E[\xi] = \int_{-\infty}^{\infty} x dF(x) \qquad V[\xi] = \int_{-\infty}^{\infty} (x - E[\xi])^2\, dF(x)$$

note that these definitions are valid for discrete and continuous random variables.

5 Stochastic Processes

5.1 Definitions

The idea now is to study sequences of random variables. A stochastic process is similar to a random variable, with the difference that it also depends on time. Adding the time dimension adds notation, but it does not change any of the main ideas. For convenience we will first go over the definition of a random variable.

Definition 5.1. (Random variable) Let $(\Omega, \mathcal{A}, P)$ be a probability space and $x : \Omega \to \mathbb{R}$ a real valued function. x is a random variable if and only if x is measurable, that is, if and only if $x^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}$, where $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{R}$. We further establish the same notation:

i. An outcome is an element $\omega \in \Omega$.

ii. An event is a measurable subset of Ω: $A \in \mathcal{A}$.

iii. The real number $x(\omega)$ is a realization of the random variable.

iv. The probability measure for x is then: $\mu(B) = P(x^{-1}(B)) = P(\{\omega \in \Omega | x(\omega) \in B\})$, for $B \in \mathcal{B}$.

v. The distribution function for f is: $G(b) = \mu((-\infty, b])$, for $b \in \mathbb{R}$.

Now we can work on adding the time dimension to the definition of a random variable. In general time can be discrete or continuous, but in what follows we will assume that time is continuous starting at 0 and going on forever, so $t \in [0, \infty)$. Intuitively a stochastic process is formed by function $x : [0, \infty) \times \Omega \to \mathbb{R}$ that gives a realization for every outcome and time. At every point in time the random variable takes a variable, the sequence of those values forms the realization (path) of the stochastic process.

The question is on how to measure the possible outcomes of the random variable through time. We need a way of determining where the random variable is at a certain point in time, and where it has been, but that does not provide information about the value of future realizations. This is achieved using a filtration.

Definition 5.2. (Filtration) Let $\mathcal{A}$ be a σ-algebra. The set $\mathbb{A} = \{\mathcal{A}_t | t \geq 0\}$ is a filtration if $\mathcal{A}_t \subseteq \mathcal{A}$ and $\mathcal{A}_s \subseteq \mathcal{A}_t$ for all $t \geq 0$ and $s \leq t$. $\mathcal{A}_t$ is the set of events known at time t.

Now we can define a stochastic process as a function that is measurable in a filtered space.

Definition 5.3. (Stochastic Process) Let $(\Omega, \mathbb{A}, P)$ be a filtered probability space with a time index $t \in \mathbb{R}_+$, and let $\mathcal{B}_+$ be the Borel sets of $\mathbb{R}_+$. A stochastic process is a function $x : [0, \infty) \times \Omega \to \mathbb{R}$ that is measurable with respect to $\mathcal{B}_+ \times \mathcal{A}$ (that is, x is jointly measurable in (t, ω)). Moreover:

i. For all $t \in \mathbb{R}_+$ and $\omega \in \Omega$, $x(t, \omega)$ is measurable with respect to $\mathcal{A}_t$, where $\mathcal{A}_t$ is in the filtration $\mathbb{A}$.

ii. For all $t \in \mathbb{R}_+$, $x(t,\cdot):\Omega \to \mathbb{R}$ is an ordinary random variable on the probability space $(\Omega, \mathcal{A}_t, P_t)$.

iii. For all $\omega \in \Omega$, $x(\cdot,\omega):\mathbb{R}_+ \to \mathbb{R}$ is a Borel measurable function. This is called the **sample path** of x.

5.2 Discrete time examples

It is not hard to come up with examples of discrete time stochastic processes. They are often used to model the behavior of many stationary economic variables by means of ARMA(p,q) representations, as well non-stationary variables usually related to random walks.

To fix ideas we start with the simple example of a (fair) coin toss. There are two possible outcomes, so $\Omega = \{H, T\}$, when tossing the coin is always possible to know which outcome occurred, and whether or not the coin was tossed, this gives: $\mathcal{A} = \{\{H\}, \{T\}, \emptyset, \Omega\}$. Finally the probability distribution P assigns values to sets in the σ-algebra $\mathcal{A}$:

$$P(\{H\}) = P(\{T\}) = \frac{1}{2} \qquad P(\emptyset) = 0 \qquad P(\Omega) = 1$$

Now we can define a random variable $\epsilon:\Omega \to \mathbb{R}$ as: $\epsilon(H) = 1$ and $\epsilon(T) = -1$. ϵ is a random variable with respect to the probability space $(\Omega, \mathcal{A}, P)$. As will be the case almost always we can dispense of the outcome space Ω for most applications and just refer to the random variable and the probability distribution induced over its values. In this way we have: $\epsilon \in \{-1, 1\}$ with $\Pr(\epsilon = 1) = \Pr(\epsilon = -1) = 1/2$.

Furthermore we can extend this example to define the stochastic process that comes up from the repeated coin toss. In this case time is discrete and finite $t \in \{1, 2, 3\}$ and at each time a coin is tossed, then the random variable variable ϵ_t is defined as the value of ϵ given the outcome of the t^{th} coin toss. The sequence $\{\epsilon_t\}_{t=1}^{3}$ is a stochastic process with respect to the filtered probability space $(\Omega, \mathbb{A}, P)$, where:

$$\Omega = \{(H,H,H), (H,H,T), (H,T,H), (H,T,T), (T,H,H), (T,T,H), (T,H,T), (T,T,T)\}$$

$$\mathcal{A} = 2^{\Omega} \qquad P(\omega) = \frac{1}{8} \quad \forall_{\omega \in \Omega}$$

The filtration is established taking into account that at each point in time only the outcome of current and past tosses is known:

$$\mathcal{A}_1 = \{\emptyset, \Omega, \{(H,H,H), (H,H,T), (H,T,H), (H,T,T)\}, \{(T,H,H), (T,T,H), (T,H,T), (T,T,T)\}\}$$

$$\mathcal{A}_2 = \{\emptyset, \Omega, \{(H,H,H), (H,H,T)\}, \{(T,H,H), (T,H,T)\}, \{(H,T,H), (H,T,T)\}, \{(T,T,H), (T,T,T)\}\}$$

$$\mathcal{A}_3 = \mathcal{A}$$

So, in the first σ-algebra all outcomes for which the first toss comes up heads are indistinguishable from each other, in the second σ-algebra one can distinguish between outcomes that have the sequence $\{H, T\}$ and $\{H, H\}$, but no information is given about the outcome of third toss. This same ideas apply if time goes on forever, so we can define our stochastic process over $t \in \mathbb{N}$.

In the previous example the stochastic process obtained satisfies the property of being iid (identically and independently distributed). The values of the stochastic process at each point in time are independent from its previous values, and they all have the same probabilities of occurring.

We now use our stochastic process $\{\epsilon_t\}$ to define a random walk. Random walks are particularly useful to understand the behavior of continuous time stochastic processes. As we will see the building block of most of them is the continuous time approximation of a random walk.

Example 5.1. (Random Walk Process) Consider a stochastic process x. Denote by x_t the value of x at time t, and fix the initial value x_0. x_t is assumed to evolve according to:

$$x_t = x_{t-1} + \epsilon_t \qquad \text{for } t \geq 1$$

ϵ_t is a random variable that can take two values $\{-1, 1\}$, and its probability distribution is independent of time, so that: $\Pr(\epsilon_t = 1) = \Pr(\epsilon_t = -1) = \frac{1}{2}$.

Note that given the starting value x_0 the variable x_t can only take on discrete values. For instance, for $x_0 = 0$ and t odd they are $\{-t, \ldots, -1, 0, 1, \ldots, t\}$, and for t even they are $\{-t, \ldots, -2, 0, 2, \ldots, t\}$. These values tell you which paths of the process cam be known at time t.

Finally, this process has no drift. Given an initial value x_0 the expected value of x_t for any t is x_0 ($E[x_t] = x_0$), this follows from the expected value of each change being $E[x_t - x_{t-1}] = E[\epsilon_t] = 0$.

This process can be generalized in many ways. The most useful one for our purposes is to allow for drift, which can be done by changing the probabilities of the random variable ϵ_t, letting $\Pr(\epsilon_t = 1) = p$ and $\Pr(\epsilon_t = -1) = 1 - p$ achieves the desired result. If $p > 1/2$ the process will have positive drift.

5.3 Brownian motion (Wiener processes)

A Brownian motion, or Weiner process, is a continuous time stochastic process $(W(t))$ that satisfies three properties:

i. $W(t)$ has continuous sample paths.

ii. $W(t)$ has stationary independent increments.

iii. Increments of $W(t)$ over a finite interval of time are normally distributed with variance that increases linearly in time.

The first property implies that a Brownian motion has no jumps, so as the time interval goes to zero the change in the process must also go to zero. The second and third properties imply that the change in $W(t)$ over some interval of length Δt must satisfy:

$$\Delta W = \epsilon_t \sqrt{\Delta t} \qquad \epsilon_t \sim N(0, 1)$$

which we write as $dW = \epsilon_t\sqrt{dt}$ as $\Delta t \to 0$. Note that this implies that:

$$E\left[dW\right] = E\left[\epsilon_t\right]\sqrt{dt} = 0 \qquad V\left[dW\right] = E\left[\epsilon_t^2\right]dt = dt$$

Moreover we assume that ϵ_t is serially uncorrelated, i.e., $E\left[\epsilon_t\epsilon_s\right] = 0$ for $t \neq s$, so the values of dW for any two different time intervals are independent.

Its easy to note the relation between the Brownian motion and the random walk processes. In discrete time we had $\Delta x_t = x_t - x_{t-1} = \epsilon_t\Delta t$, where $\Delta t = 1$. We will use this fact when approximating Brownian motions using random walks as $\Delta t \to 0$.

To see that this representation implies the third property consider a time interval that starts at t and ends at T, and divide into n intervals of length $\Delta t = {}^{T}\!/_{n}$. Then we have:

$$W\left(t+T\right) - W\left(t\right) = \sum_{i=1}^{n}\epsilon_i\sqrt{\Delta t}$$

What we want to show is that $W\left(t+T\right) - W\left(t\right) \sim N\left(0,t\right)$. To prove this we can use the Central Limit Theorem:

Theorem 5.1. ***(Central Limit Theorem)*** *If* $\{\epsilon_1, \epsilon_2, \epsilon_3, \ldots\}$ *are iid (but not necessarily normal) with* $E\left[\epsilon_i\right] = \mu < \infty$ *and* $V\left[\epsilon_i\right] = \sigma^2 < \infty$*, then* $Z_n = \sqrt{n}\frac{\sum_{i=1}^{n}\epsilon_i - n\mu}{\sigma} \to N\left(0,1\right)$ *as* $n \to \infty$*.*

Note that ϵ_i already satisfies being iid and $E\left[\epsilon_i\right] = 0$ and $V\left[\epsilon_i\right] = 1$, so $Z_n = \sqrt{n}\sum_{i=1}^{n}\epsilon_i$. Then we can write:

$$W\left(t+T\right) - W\left(t\right) = \sqrt{T}Z_n$$

By the CLT this converges to a $N\left(0,T\right)$ as $n \to \infty$.

A Brownian motion can be generalized to have drift μ and variance σ^2. This is done by adjusting the way the increments of the stochastic process work:

$$dx = \mu dt + \sigma dW$$

In this case the increments are given by a non-stochastic component μdt, which indicates that the process will drift by μ per unit of time deterministically if there are no shocks, and by a stochastic component σdW, where σ is scaling the variance of the increments of the Weiner process W. This process satisfies:

$$E\left[dx\right] = \mu dt \qquad V\left[dx\right] = \sigma^2 dt$$

5.3.1 Random walk approximation of a Brownian motion

As mentioned above we can use the similarities between the increments of a Brownian motion and the increments of a random walk to approximate continuous time processes using discrete time ones. This is important because of two reasons: it helps explain the mechanics of the continuous time model, and it provides an algorithm for simulation in the computer.

Our objective is to approximate the a Brownian motion with drift:

$$dx = \mu dt + \sigma dW$$

We will approximate with a discrete time process y whose increments are h with probability p and $-h$ with probability $1-p$. This gives:

$$E\left[\Delta y\right] = ph - (1-p)\,h = (2p-1)\,h$$

$$V\left[\Delta y\right] = E\left[(\Delta y)^2\right] - \left(E\left[\Delta y\right]\right)^2 = \left(1 - (2p-1)^2\right) h^2$$

In order to get the approximation we need to choose values for h, p and Δt so that:

$$\begin{aligned} \mu \Delta t &= (2p-1)\,h \\ \sigma^2 \Delta t &= 4p\,(1-p)\,h^2 \end{aligned}$$

Solving for p we get:

$$p^2 - p + \frac{\sigma^2}{4\left(\sigma^2 + \mu^2 \Delta t\right)} = 0$$

The roots of these equation are:

$$\begin{aligned} p &= \frac{1}{2}\left(1 \pm \sqrt{1 - \frac{\sigma^2}{\left(\sigma^2 + \mu^2 \Delta t\right)}}\right) \\ &= \frac{1}{2}\left(1 \pm \frac{\mu\sqrt{\Delta t}}{\sqrt{\sigma^2 + \mu^2 \Delta t}}\right) \\ &\approx \frac{1}{2}\left(1 \pm \frac{\mu}{\sigma}\sqrt{\Delta t}\right) \end{aligned}$$

where the approximation follows if Δt is small enough relative to σ^2/μ^2, since we are taking Δt close to zero this assumption is satisfied. We further choose only the "+" root since that way $p \geq 1/2$ when $\mu \geq 0$.

Now we can find a value for h:

$$\begin{aligned} \sigma^2 \Delta t &= 4p\,(1-p)\,h^2 \\ \sigma^2 \Delta t &= 2\left(1 + \frac{\mu}{\sigma}\sqrt{\Delta t}\right)\left(1 - \frac{1}{2}\left(1 + \frac{\mu}{\sigma}\sqrt{\Delta t}\right)\right) h^2 \\ \sigma^2 \Delta t &= \left(1 + \frac{\mu}{\sigma}\sqrt{\Delta t}\right)\left(1 - \frac{\mu}{\sigma}\sqrt{\Delta t}\right) h^2 \\ \sigma^2 \Delta t &= \left(1 - \left(\frac{\mu}{\sigma}\right)^2 \Delta t\right) h^2 \\ \sigma^2 \Delta t &\approx h^2 \\ \sigma\sqrt{\Delta t} &\approx h \end{aligned}$$

As before we can disregard the term $\left(\frac{\mu}{\sigma}\right)^2 \Delta t$ as long as Δt is small enough relative to σ^2/μ^2.

As an exercise can verify that the first equation also holds:

$$\begin{aligned}
\mu\Delta t &= (2p-1)\,h \\
\mu\Delta t &= (2p-1)\,\sigma\sqrt{\Delta t} \\
\frac{\mu}{\sigma}\sqrt{\Delta t} &= 2p-1 \\
\frac{1}{2}\left(1+\frac{\mu}{\sigma}\sqrt{\Delta t}\right) &= p
\end{aligned}$$

In order to simulate a Brownian motion with parameters (μ, σ) we can do as follows:

i. Set a Δt small relative to $\frac{\sigma^2}{\mu^2}$.

ii. Set $p = \frac{1}{2}\left(1+\frac{\mu}{\sigma}\sqrt{\Delta t}\right)$ and $h = \sigma\sqrt{\Delta t}$.

iii. Simulate the increments of x by drawing realization of a random variable ϵ_t that takes value h with probability p and $-h$ with probability $1-p$.

5.4 Ito processes

Ito processes are the generalization of Brownian motions. Their drift and variance is allowed to depend on the level of the process and the time:

$$dx = \mu(x,t)\,dt + \sigma(x,t)\,dW \tag{5.1}$$

where the functions μ and σ give the value of the mean and standard deviations of the increments of the process x:

$$\begin{aligned}
\mu(x,t) &= \lim_{\Delta\to 0^+}\frac{1}{\Delta}E\left[x(t+\Delta)-x(t)\,|x(t)=x\right] \\
(\sigma(x,t))^2 &= \lim_{\Delta\to 0^+}\frac{1}{\Delta}E\left[(x(t+\Delta)-x(t))^2\,|x(t)=x\right]
\end{aligned}$$

For future reference note that an Ito process can also be represented as:

$$x(t) = x(0) + \int_0^t \mu(x,s)\,ds + \int_0^t \sigma(x,s)\,dW(s)$$

where the last term is a stochastic integral. Stochastic integrals play an important role in the theory of stochastic processes, for now it suffices to state the following result.

Proposition 5.1. *Let $x(t)$ be an integrable function, then $E\left[\int_0^t x(s)\,dW(s)\right] = 0$.*

This proposition states that the expected value of a stochastic integral is identically zero. The derivation of the result, along with other properties can be found in Stokey (2009, Sec. 3.2).

Two Ito process are of particular importance. They are presented in the examples below.

Example 5.2. (Geometric Brownian notion) A Geometric Brownian motion is an Ito process with $\mu(x,t) = \mu x$ and $\sigma(x,t) = \sigma x$, so:

$$dx = \mu x dt + \sigma x dW$$

A geometric Brownian motion can be thought of as a Brownian motion where the properties apply to percentage increments instead of increments, note that:

$$\frac{dx}{x} = \mu dt + \sigma dW$$

So the percentage increment, ${}^{dx}\!/_{x}$, are normally distributed with mean $\mu \Delta t$ and variance $\sigma^2 \Delta t$.

Example 5.3. (Ornstein-Uhlenbeck process) Unlike the previous processes an OU process is mean reverting, similar to an AR(1) process in discrete time. An OU process is an Ito process with $\mu(x,t) = \mu(\overline{x} - x)$ and $\sigma(x,t) = \sigma$. Note that if $x > \overline{x}$ then the process drifts down, and if $x < \overline{x}$ the process drifts up.

$$dx = \mu(\overline{x} - x) dt + \sigma dW$$

5.5 Jump processes - Poisson Processes

Jump processes are a type of stochastic process that has discontinuous paths. Jump process change by discrete amounts when a certain outcome occurs. The most important Jump process is the Poisson process, which is just a jump process such that the time of the jumps follows a Poisson distribution. To define it let λ be the mean arrival rate of a jump and u the size of the change of the process (usually $u = 1$, but in general u can be itself a random variable). Then for some process q we have:

$$dq = \begin{cases} 0 & \text{with prob. } 1 - \lambda dt \\ u & \text{with prob. } \lambda dt \end{cases}$$

We can now define a more general process that depends on the Jump process q:

$$dx = f(x,t) dt + g(x,t) dq \tag{5.2}$$

where absent a jump x evolves deterministically according to the function f, and when there is a jump it moves according to function g. Note that:

$$E[dx] = f(x,t) dt + \lambda E_u [g(x,t) u] dt$$

Part II
Stochastic Calculus

This part of the course develops the mathematical tools necessary to study how random variables affect optimization problems. The most important result is Ito's Lemma, which defines the way in which we can take derivatives of functions that depend on diffusions. Then we can apply Ito's Lemma to problems of dynamic optimization, with special attention to stopping time problems. Finally we apply it to the characterization of the distribution of a random variable. This is done by means of the Kolmogorov forward equation.

All these sections follow closely Dixit and Pindyck (1994), with some portions adapted from Stokey (2009).

6 Ito's Lemma

We are often concerned with the behavior of functions of stochastic processes, in particular the differentials of those functions. The number one example at hand is to know how the value of an asset (or an option) evolves over time. Ito's Lemma gives a way to compute those differentials. This relates the functions we are interested in to the stochastic differential equation that governs the underlying stochastic process.

Consider a function $F(x,t)$ that depends on a stochastic process x. x is assumed to be an Ito process following:

$$dx = \mu(x,t)\,dt + \sigma(x,t)\,dW \tag{6.1}$$

Normal calculus rules would give the differential of F as:

$$dF = \frac{\partial F}{\partial x}dx + \frac{\partial F}{\partial t}dt$$

Although not always clear, one of the reasons for expressing the differential without resorting to higher order terms is that those terms depend on dt^2, $dt^3 \ldots$. As $dt \to 0$ all higher order terms go to zero faster, and are hence ignored. But stochastic process add a new factor since their components depend of time through $\sqrt{dt}$, so square terms like $(dx)^2$ must also be considered.

A second order Taylor expansion of F gives:

$$dF = \frac{\partial F}{\partial x}dx + \frac{\partial F}{\partial t}dt + \frac{1}{2}\left(\frac{\partial^2 F}{\partial x^2}(dx)^2 + \frac{\partial^2 F}{\partial t^2}(dt)^2 + \frac{\partial^2 F}{\partial t^2}(dx)(dt)\right)$$

As shown in Øksendal (2003, Sec. 4.1) $dW dt = dt^2 = 0$, they can be safely ignored since they depend on terms of order higher than dt. That leaves us with:

$$dF = \frac{\partial F}{\partial x}dx + \frac{\partial F}{\partial t}dt + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}(dx)^2 \tag{6.2}$$

From the definition of our Ito process we get:

$$\begin{aligned}(dx)^2 &= \left(\mu^2(x,t)(dt)^2 + 2\mu(x,t)\,\sigma(x,t)\,dt dW + \sigma^2(x,t)(dW)^2\right)\\ &= \sigma^2(x,t)(dW)^2\end{aligned}$$

We can again drop the terms involving $(dt)^2$ and $(dt dW)$, and also show that $(dW)^2 = dt$ (recall that $E\left[(dW)^2\right] = dt$). The proof is not hard and can be found in Øksendal (2003, Sec. 4.1). Replacing:

$$\begin{aligned}dF &= \left(\frac{\partial F}{\partial x}dx + \frac{\partial F}{\partial t}dt\right) + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}\left(\sigma^2(x,t)\,dt\right)\\ dF &= \left(\frac{\partial F}{\partial t} + \mu(x,t)\frac{\partial F}{\partial x} + \frac{1}{2}\sigma^2(x,t)\frac{\partial^2 F}{\partial x^2}\right)dt + \sigma(x,t)\frac{\partial F}{\partial x}dW \qquad (6.3)\end{aligned}$$

This derivation (Ito's Formula) means that $y = F(x,t)$ is itself an Ito process with $\mu_y(x,t) = \left(F_t + \mu(x,t)F_x + \frac{1}{2}\sigma^2(x,t)F_{xx}\right)$ and $\sigma_y(x,t) = \sigma(x,t)F_x$ as parameters. Unsurprisingly, the expected value of y is $\mu_y(x,t)$ and its variance is $\sigma_y(x,t)\,dt$.

6.1 Application to geometric brownian motion

We can use Ito's Lemma to obtain the properties of different stochastic processes. For instance the Geometric Brownian motion can be shown to be the exponential of a standard brownian motion, or equivalently it can be shown that the logarithm of a geometric brownian motion is a brownian motion.

Let x be a geometric brownian motion satisfying:

$$dx = \mu x dt + \sigma x dW$$

and $y = \ln x$. By Ito's Lemma:

$$\begin{aligned} dy &= \left(\mu x \cdot \frac{1}{x} + \frac{1}{2}\sigma^2 x^2 \cdot \left(\frac{-1}{x^2}\right)\right) dt + \sigma x \cdot \frac{1}{x} dW \\ &= \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma dW \end{aligned}$$

thus y is a brownian motion with parameters $\mu_y = \mu - \frac{1}{2}\sigma^2$ and $\sigma_y = \sigma$. Note that the drift of y is lower than the drift of x, since the logarithm is a concave function Jensen's inequality implies that the expected value of the log is lower.

We can also obtain the expected value of x by noting that:

$$x(t) = x(0) + \int_0^t \mu x(s)\, ds + \int_0^t \sigma x(s)\, dW(s)$$

taking expectations gives:

$$E[x(t)] = x(0) + \int_0^t \mu E[x(s)]\, ds$$

recalling that the third term is a stochastic integral, and hence has expected value equal to zero. From this equation we can derive a first order differential equation for the expected value of x:

$$dE[x] = \mu E[x]\, dx$$

The solution for this equation, given the boundary condition $E[x(0)] = x(0)$ is:

$$E[x] = x(0)\, e^{\mu t}$$

Finding the variance (and other moments) works in the same way. For the variance we want to obtain an expression for x^2, so first consider the function $f(x) = x^2$. By Ito's Lemma we get:

$$\begin{aligned} df &= \left(2\mu x^2 + \sigma^2 x^2\right) dt + 2\sigma x^2 dW \\ x^2 = f(x) &= x_0^2 + \left(2\mu + \sigma^2\right) \int_0^t x^2(s)\, ds + 2\sigma \int_0^t x^2(s)\, dW(s) \end{aligned}$$

We can now take expectations to obtain:

$$E\left[x^2\right] = x_0^2 + \left(2\mu + \sigma^2\right) \int_0^t E\left[x^2(s)\right] ds$$

which leads to a differential equation for $E\left[x^2\right]$:

$$
\begin{aligned}
dE\left[x^2\right] &= \left(2\mu+\sigma^2\right)E\left[x^2\right] \\
E\left[x^2\right] &= x^2(0)\, e^{\left(2\mu+\sigma^2\right)t}
\end{aligned}
$$

the variance is then:

$$
\begin{aligned}
V[x] &= E\left[x^2\right]-E[x]^2 \\
&= x^2(0)\, e^{\left(2\mu+\sigma^2\right)t}-x^2(0)\, e^{2\mu t} \\
&= x^2(0)\, e^{2\mu t}\left(e^{\sigma^2 t}-1\right)
\end{aligned}
$$

Some applications are shown below:

Example 6.1. Consider an asset that gives flow payoffs x that evolve according to a geometric brownian motion

$$
dx=\mu x dt+\sigma x dW
$$

we can compute the expected discounted value of holding that asset easily using the results above:

$$
E\left[\int_0^\infty e^{-\rho t}x(t)\,dt\right]=\int_0^\infty e^{-\rho t}E[x(t)]\,dt=\int_0^\infty x(0)\,e^{-(\rho-\mu)t}dt=\frac{x_0}{\rho-\mu}
$$

Example 6.2. Now consider an agent that receives flow consumption of x, which evolves again as a geometric brownian motion. The agent's utility is CRRA, so that $u(x)=\frac{x^{1-\theta}}{1-\theta}$. We want to know the expected present value of utility.

$$
E\left[\int e^{-\rho t}u(x(t))\,dt\right]=\int e^{-\rho t}E[u(x(t))]\,dt
$$

To know it we need to compute $E[u(x(t))]$. From Ito's Lemma we have:

$$
\begin{aligned}
du &= \left(\mu x\cdot x^{-\theta}+\frac{1}{2}\sigma^2x^2\cdot\left(-\theta x^{-\theta-1}\right)\right)dt+\sigma x\cdot x^{-\theta}dW \\
du &= (1-\theta)\left(\mu-\frac{\theta}{2}\sigma^2\right)\frac{x^{1-\theta}}{1-\theta}dt+(1-\theta)\,\sigma\frac{x^{1-\theta}}{1-\theta}dW \\
du &= (1-\theta)\left(\mu-\frac{\theta}{2}\sigma^2\right)udt+(1-\theta)\,\sigma udW
\end{aligned}
$$

Thus, u is itself a geometric brownian motion (actually if x is a brownian motion x^k is a geometric brownian motion). Using our previous results we have:

$$
E[u]=u(x(0))\,e^{(1-\theta)\left(\mu-\frac{\theta}{2}\sigma^2\right)t}
$$

So we have:

$$
E\left[\int e^{-\rho t}u(x(t))\,dt\right]=\int e^{-\rho t}E[u(x(t))]\,dt=\frac{(x(0))^{1-\theta}}{(1-\theta)\left(\rho-(1-\theta)\left(\mu-\frac{\theta}{2}\sigma^2\right)\right)}
$$

6.2 Poisson Processes

Similar, and simpler, results can be obtained if x follows a Poisson process:

$$dx = f(x,t)\,dt + g(x,t)\,dq$$

and we have a function $H(x,t)$ that depends on x. Unlike the Ito Process the Poisson process does not depend on $\sqrt{dt}$, so higher order terms in the Taylor expansion can be ignored altogether to get:

$$\begin{aligned} dH &= H_t dt + H_x dx \\ &= (H_t + f(x,t)\,H_x)\,dt + g(x,t)\,H_x dq \end{aligned}$$

The expected value of this change must take into account the probability of a jump in q (given by λdt), so we have:

$$E[dH] = (H_t + f(x,t)\,H_x)\,dt + \lambda E_u\left[H(x + ug(x,t),t) - H(x,t)\right]dt \tag{6.4}$$

it follows, by using the identity function that $E[dx] = f(x,t)\,dt + \lambda E_u[ug(x,t)]\,dt$.

We can apply this result to a couple examples taken from Dixit and Pindyck (1994):

Example 6.3. Consider an individual that lives forever and receives a wage $w(t)$ at each point in time. The wage increases by ϵ at random times, following a Poisson process with arrival rate λ, so:

$$dw = \epsilon dq$$

The individual wants to know the expected discounted value of taking the job we need to compute:

$$V(w) = E\left[\int_0^\infty e^{-\rho t} w(t)\,dt\right]$$

The function V (a value function) has a recursive representation, this is easier to see in the discrete time approximation. Consider a period of length Δt, then:

$$\begin{aligned} V(w(t)) &= w(t)\,\Delta t + \frac{1}{1+\rho\Delta t} E\left[V(w(t+\Delta t))\right] \\ (1+\rho\Delta t)\,V(w(t)) &= (1+\rho\Delta t)\,w(t)\,\Delta t + E\left[V(w(t+\Delta t))\right] \\ \rho(\Delta t)\,V(w(t)) &= (1+\rho\Delta t)\,w(t)\,\Delta t + E\left[V(w(t+\Delta t)) - V(w(t))\right] \\ \rho V(w(t)) &= (1+\rho\Delta t)\,w(t) + \frac{E[\Delta V]}{\Delta t} \end{aligned}$$

Taking the limit as $\Delta t \to 0$ we get:

$$\rho V = w + \frac{E[dV]}{dt} \tag{6.5}$$

Staying in the job works just like an asset, with a normal return at rate ρ being equal to the sum of the dividend (in this case given by the wage) and the expected capital gains (from changes in the wage). In the expression above $\frac{E[dV]}{dt} = \lim_{\Delta t \to 0} \frac{1}{\Delta t} E[\Delta V]$

We can apply the formula from above:

$$E\left[dH\right] = \left(H_t + f\left(x,t\right)H_x\right)dt + \lambda E_u\left[H\left(x + ug\left(x,t\right),t\right) - H\left(x,t\right)\right]dt$$

where $H = V$, $x = w$, $f\left(x,t\right) = 0$ and $g\left(x,t\right) = \epsilon$ and $u = 1$ with certainty:

$$\begin{aligned} E\left[dV\right] &= \lambda\left(V\left(w+\epsilon\right) - V\left(w\right)\right)dt \\ &= \lambda\epsilon\left(\int_0^\infty e^{-\rho t}dt\right)dt \\ &= \frac{\lambda\epsilon}{\rho}dt \end{aligned}$$

This leaves us with an explicit solution for V:

$$V = \frac{w}{\rho} + \frac{\lambda\epsilon}{\rho^2}$$

V is equal to an asset that pays the current wage forever plus the capitalized value of the average raise in wages per unit of time.

Example 6.4. Consider now a firm that produces using capital. As long as capital is operational a flow profit of π is obtained, but capital becomes obsolete when new technologies arrive. These innovations occur at random times following a Poisson process with arrival rate λ. Once the innovation arrives and the capital becomes obsolete the firm goes out of business forever.

The value of the firm follows a process:

$$dV = -Vdq$$

The return can be found as before:

$$\rho V = \pi + \frac{1}{dt}E\left[dV\right]$$

To find $E\left[dV\right]$ we can again use our formula with $H = V$, the identity function:

$$E\left[dV\right] = -\lambda V dt$$

replacing we get:

$$\begin{aligned} \rho V &= \pi - \lambda V \\ V &= \frac{\pi}{\rho + \lambda} \end{aligned}$$

Note that this is equivalent to solving:

$$V = \int_0^\infty e^{-(\rho+\lambda)t}\pi dt$$

This should not be a surprise. Consider the case where there are no shocks and the firm can operate forever with certainty. Then V is:

$$V = \int_0^\infty e^{-\rho t} \pi dt = \frac{\pi}{\rho}$$

Now the firm shuts down with a certain probability, given by the arrival of the Poisson shock. Then:

$$V = E\left[\int_0^\infty e^{-\rho t} \pi dt\right] = \int_0^\infty \Pr\left[\text{No shock until time } t\right] e^{-\rho t} \pi dt$$

The probability of there being no shocks is known:

$$\Pr\left[\text{No shock until time } t\right] = e^{-\lambda t}$$

Replacing gives the desired result.

7 Dynamic Programming

In dynamic programming we aim to develop tools for solving problems that involve actions through time, that in turn affect the total value obtained by the agent takin the decisions. The key of dynamic programming is that it focuses on the current decision being taken and its effect on the continuation value for the agent, rather than try to solve for the whole sequence of actions at once.

7.1 Discrete time overview

Dynamic Programming

To build up to the concepts of dynamic programming in continuous time we will first consider a simple discrete time problem of a firm that must invest a fixed amount I to set up the operation of the firm. Once the firm is operational the firm produces one unit of good every period.The current price of the good is known and given by p_0, in the second period the price can go up or down:

$$p_1 = \begin{cases} (1+u)\,p_0 & \text{with prob. } q \\ (1-d)\,p_0 & \text{with prob. } 1-q \end{cases}$$

After that the price is constant. Hence, the firm's decision is whether to invest in the first period, in the second, or not to invest at all. It makes no sense to wait any longer since no new information will arrive after the initial change in price. The firm discounts future payments with an interest rate r.

We can solve the problem by tracing the decisions that the firm can take. In the second period, once the price is known and assuming that the firm is not yet in operation, the firm can either invest or not. If the firm does not invest it gets zero payoff, if it invests it gets:

$$F_1(p_1) = p_1 + \frac{p_1}{1+r} + \frac{p_1}{(1+r)^2} = p_1 \sum_{i=0}^{\infty} \frac{1}{(1+r)^i} = \frac{1+r}{r} p_1$$

The payoff of the firm is then:

$$V_1(p_1) = \max\{F_1(p_1) - 0, 0\}$$

Knowing this is relevant because if the firm does not invest in the first period it can always do so later, so V_1 constitutes the continuation payoff of the firm. The payoff to the firm if it does not invest in the first period is then:

$$\frac{1}{1+r} E[V_1(p_1)] = \frac{1}{1+r}\left(qV_1((1+u)\,p_0) + (1-q)\,V_1((1-d)\,p_0)\right)$$

If the firm invests in the first period the payoff is:

$$
\begin{aligned}
F_0(p_0) &= p_0 + \frac{1}{1+r} E\left[F_1(p_1)\right] \\
&= p_0 + \frac{1}{1+r}\left(qF_1\left((1+u)\,p_0\right) + (1-q)\,F_1\left((1-d)\,p_0\right)\right) \\
&= p_0 + \left(\frac{q}{r}(1+u)\,p_0 + \frac{1-q}{r}(1-d)\,p_0\right) \\
&= \frac{1}{r}\left(1 + r + q\,(u+d) - d\right) p_0
\end{aligned}
$$

So, the value of the firm is:

$$
V_0(p_0) = \max\left\{F_0(p_0) - I, \frac{1}{1+r} E\left[V_1(p_1)\right]\right\}
$$

In this example we already see the basics of dynamics programming, splitting the problem into the decision at hand (invest or not invest) and the continuation value that they entail. The example also highlights one of the recurring topics of the course: option value. The firm has an option that allows it to invest any of the two dates. Waiting in this problem has value, since investing in the future also means to invest with better information. In fact we can compute the value of this option (to wait) by comparing the value that the firm would have if it was forced to take a decision in the first period:

$$
\Omega_0(p_0) = \max\left\{F_0(p_0) - I, 0\right\}
$$

with the value that includes the possibility of action in the second period:

$$
V_0(p_0) - \Omega_0(p_0)
$$

We can now extend this simple model to allow for action in many periods (more than two). Consider a firm that operates for $T < \infty$ periods. In each period the firm will choose the value of a control variable u that affects (potentially) the per-period payoffs of the firm, namely the profits, and the evolution of a random variable x. x is assumed to follow a Markov process so that the CDF of x_{t+1} is $\Phi_t\left(x_{t+1}|x_t, u_t\right)$. The random variable x is also allowed to affect payoffs, so per-period payoffs are: $\pi\left(u_t, x_t\right)$.

The firm discounts the future at rate $\frac{1}{1+\rho}$ and receives a final payoff pf $\Omega_T\left(x_T\right)$ in the last period. The objective is:

$$
V_0(x_0) = \max_{\{u_t\}_{t=0}^{T-1}} E\left[\sum_{t=0}^{T-1}\left(\frac{1}{1+\rho}\right)^t \pi\left(u_t, x_t\right) + \left(\frac{1}{1+\rho}\right)^T \Omega_T\left(x_T\right)\right] \tag{7.1}
$$

dynamic programming allows us to write the problem recursively. In the last period we have:

$$
V_{T-1}\left(x_{T-1}\right) = \max_{u_{T-1}} \pi\left(u_{T-1}, x_{T-1}\right) + \left(\frac{1}{1+\rho}\right) E\left[\Omega_T\left(x_T\right)|x_{T-1}, u_{T-1}\right] \tag{7.2}
$$

For all other periods we can use the notion of continuation payoffs to obtain:

$$V_t(x_t) = \max_{u_t} \pi(u_t, x_t) + \left(\frac{1}{1+\rho}\right) E\left[V_{t+1}(x_{t+1}) | x_t, u_t\right] \tag{7.3}$$

The problem can then be solved by backwards induction, choosing contingent plans for $u_t(x_t)$ one period at a time, instead of tackling the more complicated problem of choosing the whole sequence of $\{u_t\}$.

When time is not finite, there is no terminal date, and we cannot use backwards induction to solve the problem. In this case the value of the firm itself is also independent of time, since each period is just like the next. We then have:

$$V(x) = \max_{u} \pi(u, x) + \left(\frac{1}{1+\rho}\right) E\left[V\left(x'\right) | x, u\right] \tag{7.4}$$

the problem is now to find a function V that satisfies the equation above. The details behind the solution to this problem can be found in Stokey et al. (1989).

This setup is very versatile and can be applied to firm problems as the one above, but it is also at the core of modern macroeconomic theory. The following examples make this point in a non-stochastic version of the model.

Example 7.1. Consider an economy in which the representative consumer lives forever. There is a good in each period that can be consumed or saved as capital as well as labor. The consumer's utility function is

$$V\left(\overline{k}_0\right) = \sum_{t=0}^{\infty} \beta^t \log c_t$$

Here $0 < \beta < 1$. The consumer is endowed with 1 unit of labor in each period and with $\overline{k}_0$ units of capital in period 0. Capital fully depreciates each period. Feasible allocations satisfy

$$c_t + k_{t+1} \leq \theta k_t^{\alpha} l_t^{1-\alpha}$$

Here $\theta > 0$ and $0 < \alpha < 1$. We can formulate the problem of maximizing the representative consumer's utility subject to feasibility conditions as a dynamic programming problem. The appropriate Bellman's equation is:

$$\begin{aligned}
V(k) &= \max_{c,k',l} \{\log c + \beta V(k')\} \\
\text{s.t. } & c + k' \leq \theta k^{\alpha} l^{1-\alpha} \\
& c, k' \geq 0 \\
& 0 \leq l \leq 1
\end{aligned}$$

To solve it we guess that the value function has the form $a_0 + a_1 \log k$ and solve for the decisions of the consumer. The constraint will hold with equality because the utility function is strictly increasing in consumption, also production increases with labor and there is no disutility of it, hence there is a corner solution for labor indicating $l = 1$, so with the guess the problem becomes

$$a_0 + a_1 \log k = \max_{k' \in [0, \theta k^\alpha l^{1-\alpha}]} \log \left(\theta k^\alpha l^{1-\alpha} - k'\right) + \beta \left(a_0 + a_1 \log k'\right)$$

Then the FOC is

$$\frac{1}{\theta k^\alpha l^{1-\alpha} - k'} = \frac{\beta a_1}{k'}$$

solving for k'

$$\begin{aligned} k' &= \beta a_1 \left(\theta k^\alpha l^{1-\alpha} - k'\right) \\ &= \frac{\beta a_1 \left(\theta k^\alpha l^{1-\alpha}\right)}{1 + \beta a_1} \end{aligned}$$

Then plugging this back into the value function you get

$$a_0 + a_1 \log k = \log \left(\theta k^\alpha l^{1-\alpha} - \frac{\beta a_1 \left(\theta k^\alpha l^{1-\alpha}\right)}{1 + \beta a_1}\right) + \beta \left(a_0 + a_1 \log \left(\frac{\beta a_1 \left(\theta k^\alpha l^{1-\alpha}\right)}{1 + \beta a_1}\right)\right)$$

Collection terms with k you get

$$\begin{aligned} a_1 \log k &= \alpha \log k + \beta a_1 \alpha \log k \\ a_1 \left(\log k - \beta \alpha \log k\right) &= \alpha \log k \\ a_1 &= \frac{\alpha}{1 - \beta \alpha} \end{aligned}$$

which means the policy function is

$$\begin{aligned} k' &= \frac{\beta \frac{\alpha}{1-\beta\alpha} \left(\theta k^\alpha l^{1-\alpha}\right)}{1 + \beta \frac{\alpha}{1-\beta\alpha}} = \beta \alpha \theta k^\alpha l^{1-\alpha} \\ l &= 1 \\ c &= \theta k^\alpha l^{1-\alpha} - \beta \alpha \theta k^\alpha l^{1-\alpha} \end{aligned}$$

Optimal Stopping Time

There is another type of problem that deserves special treatment. Optimal stopping time problems are at the core of the continuous time applications in the rest of the course. In these problems the agent faces a binary choice (instead of a continuous choice as in the example above), they resemble the example of the firm at the beginning of the Section where the firm has to choose whether or not to invest. This problems are characterized by the inaction of the agent, since the agent usually acts just once, and most of the time the optimal choice is to do nothing. To characterize these problems let $\Omega(x)$ be the termination payoff received once the action is taken (and time is stopped). It depends on the value of state x. The Bellman equation is now:

$$V(x) = \max \left\{ \Omega(x), \max_u \pi(u, x) + \left(\frac{1}{1+\rho}\right) E\left[V\left(x'\right) | x, u\right] \right\} \tag{7.5}$$

We can now define a stopping time as a random variable that signals the decision to stop and take the termination payoff $\Omega(x)$. So:

$$T^{\star}=\left\{x|\Omega(x)\geq\max_{u}\pi(u,x)+\left(\frac{1}{1+\rho}\right)E\left[V\left(x^{'}\right)|x,u\right]\right\} \tag{7.6}$$

In general $T^{\star}$ can take many forms, but in most (if not all) of the relevant economic applications it will take the form: $T^{\star}=[\overline{x},\infty)$, $T^{\star}=(-\infty,\underline{x}]$ or $T^{\star}=(-\infty,\underline{x}]\cup[\overline{x},\infty)$. As an example we apply these ideas to the problem of search and unemployment, the McCall search model.

Example 7.2. Consider the following infinite horizon model. An agent searches for a job. Each period the agent receives a wage offer from a distribution $F(w)$ with bounded support $W=\left[0,\overline{W}\right]$. If accepted the agent will remain employed at that wage forever. If rejected the worker receives unemployment benefits b. Wage offers are iid over time. The worker preferences are $\sum\beta^{t}c_{t}$. Assume no borrowing or lending.

We first set up the workers decision as a dynamic programming problem:

$$\begin{aligned}V^{E}(w)&=\frac{w}{1-\beta}\\ V^{U}&=b+\beta\int\max\left\{V^{E}(\tilde{w}),V^{U}\right\}dF(\tilde{w})\end{aligned}$$

The decision of a worker when facing a wage offer w is to accept it or reject it, the worker will accept if $V^{E}(w)>V^{u}$ and reject otherwise. Then the value of the worker is:

$$\begin{aligned}V(w)&=\max\left[V^{E}(w),V^{U}\right]\\ V(w)&=\max\left[\frac{w}{1-\beta},b+\beta\int V(\tilde{w})\,dF(\tilde{w})\right]\end{aligned}$$

Now we need to show that the decision to take action (accept a job offer) is given by $T^{\star}=[\overline{w},\infty)$, where $\overline{w}$ is the reservation wage. To show this note that V^{U} is independent of the wage and that V^{E} is increasing in wages. The reservation wage satisfies:

$$\frac{\overline{w}}{1-\beta}=b+\beta\int V(\tilde{w})\,dF(\tilde{w})$$

This implies that V is constant for $w<\overline{w}$, since the offers are rejected, and it is equal to V^{E} for $w\geq\overline{w}$:

$$V(w)=\begin{cases}\frac{\overline{w}}{1-\beta} & \text{if } w<\overline{w}\\ \frac{w}{1-\beta} & \text{if } w\geq\overline{w}\end{cases}$$

It is left to find $\overline{w}$. To do this we should first solve for V^U:

$$\begin{aligned}
V^U &= b + \beta \int \max\left\{V^E(\tilde{w}), V^U\right\} dF(\tilde{w}) \\
&= b + \beta \int_0^{\overline{w}} \frac{\overline{w}}{1-\beta} dF(\tilde{w}) + \beta \int_{\overline{w}}^{\overline{W}} \frac{\tilde{w}}{1-\beta} dF(\tilde{w}) \\
&= b + \frac{\beta}{1-\beta}\left(\int_0^{\overline{w}} \overline{w} dF(\tilde{w}) + \int_{\overline{w}}^{\overline{W}} \tilde{w} dF(\tilde{w})\right) \\
&= b + \frac{\beta}{1-\beta}\left(\overline{w} - \int_{\overline{w}}^{\overline{W}} \overline{w} dF(\tilde{w}) + \int_{\overline{w}}^{\overline{W}} \tilde{w} dF(\tilde{w})\right) \\
&= b + \frac{\beta}{1-\beta}\left(\overline{w} + \int_{\overline{w}}^{\overline{W}} (\tilde{w} - \overline{w}) dF(\tilde{w})\right)
\end{aligned}$$

Note that the agent knows she is guaranteed to have $\overline{w}$ forever, finding a job just adds to the value with the wage in excess of $\overline{w}$.

Turning back to determining $\overline{w}$ we can replace V^U to get:

$$\overline{w} = b + \frac{\beta}{1-\beta} \int_{\overline{w}}^{\overline{W}} (\tilde{w} - \overline{w}) dF(\tilde{w})$$

This equation is guaranteed to have a solution for $\overline{w} \in \left[c, \overline{W}\right]$. The LHS is increasing in $\overline{w}$, while the RHS is decreasing in $\overline{w}$.

7.2 Continuous time dynamic programming

We can now turn to develop a general framework to solve dynamic problems in continuous time. To start consider the problem developed in the previous section with periods of length Δt. The agent receives a payoff $\pi(u, x)\Delta t$ every period (where π is the payoff flow), and discounts the future at a rate ρ per unit of time, so the effective discount rate for the period of length Δt is: $\frac{1}{1-\rho\Delta t}$. This leads to the following Bellman-type equation

$$V(x) = \max_u \pi(u, x)\Delta t + \left(\frac{1}{1+\rho\Delta t}\right) E\left[V\left(x'\right)|x, u\right] \tag{7.7}$$

Rearranging we get:

$$\rho V(x) = \max_u (1+\rho\Delta t)\pi(u, x) + \frac{E\left[\left(V\left(x'\right) - V(x)\right)|x, u\right]}{\Delta t} \tag{7.8}$$

Taking the limit as $\Delta t \to 0$ we get our continuous time Bellman equation:

$$\rho V(x) = \max_u \pi(u, x) + \frac{1}{dt} E\left[dV(x)|x, u\right] \tag{7.9}$$

where

$$\frac{E[dV]}{dt} = \lim_{\Delta t \to 0} \frac{1}{\Delta t} E[\Delta V]$$

Equation (7.9) works just like a non-arbitrage condition. We can thing of the agent as holding an asset with value V. The LHS gives the normal rate of return per unit time that the agent requires to hold the asset, given the discount rate ρ. The RHS gives the effective payoff of the asset, composed by the immediate flow payoff π, and the expected capital gains (brought up by changes in the value of the asset).

We can further characterize the problem given knowledge of the stochastic process that x follows. This will allow us to evaluate the expectation in (7.9). If x follows an Ito process, as in equation (5.1), then Ito's Lemma gives the following result:

$$\begin{aligned} dV &= \left(\mu(x,t)V' + \frac{1}{2}\sigma^2(x,t)V''\right)dt + \sigma(x,t)V'dW \\ E[dV] &= \left(\mu(x,t)V' + \frac{1}{2}\sigma^2(x,t)V''\right)dt \end{aligned}$$

Replacing we get the Hamilton-Jacobi-Bellman equation:

$$\rho V(x) = \max_u \pi(u,x) + \mu(x,t)V'(x) + \frac{1}{2}\sigma^2(x,t)V''(x) \tag{7.10}$$

We can take FOC with respect to u and then get a differential equation for V that we can solve.

If x follows a Poisson process, like the one in equation (5.2), we can obtain a similar result. From equation (6.4) we can compute $E[dV]$:

$$E[dV] = \left(f(x,t)V'(x)\right)dt + \lambda E_u\left[V(x+ug(x,t)) - V(x)\right]dt$$

Optimal Stopping Time and the Smooth Pasting Condition

We now go back to the stopping time problem reviewed in Section 7.1. Consider then the problem of an agent that is engaged in some activity (say running a firm). The agent gets a flow payoff of $\pi(x)$ if she continues with the activity, and $\Omega(x,t)$ if she quits the activity (stops). The value of the agent is:

$$V(x,t) = \max\left\{\Omega(x,t), \pi(x)\Delta t + \frac{1}{1+\rho\Delta t}E\left[V(x+dx,t+\Delta t)\right]\right\} \tag{7.11}$$

where x follows a diffusion process and in equation (5.1). We assume that Ω is continuous and weakly increasing in x.

In order to solve the problem we need to find regions of x where it is best for the agent to continue and those for which it is best to stop. If x is in the continuation region then:

$$V(x) = \pi(x)\Delta t + \frac{1}{1+\rho\Delta t}E\left[V\left(x'\right)\right]$$

From above we know that this implies that for x in the continuation region we have (by applying Ito's lemma):

$$\rho V(x,t) = \pi(x) + V_t(x,t) + \mu(x,t)V_x(x,t) + \frac{1}{2}\sigma^2(x,t)V_{xx}(x,t)$$

For simplicity we assume now that the continuation region has the form $x \geq x^{\star}(t)$. It is only for $x \geq x^{\star}(t)$ that the equation above holds. In order to solve it we need to impose certain boundary conditions.

By assumption we know that $V(x) = \Omega(x,t)$ for $x < x^{\star}(t)$, then, by continuity we can impose that:

$$V(x^{\star}(t), t) = \Omega(x^{\star}(t), t)$$

this is called "value-matching". Note that continuity at $x^{\star}$ is actually necessary for a solution. Suppose for a contradiction that it is optimal to stop for $x < x^{\star}(t)$, but that $V(x^{\star}(t), t) < \Omega(x^{\star}(t), t)$, since V has to be continuous in the domain $x \geq x^{\star}(t)$ (because it is the solution to a differential equation), and Ω is continuous by definition, then it holds that for x to the right of $x^{\star}(t)$, but sufficiently close to $x^{\star}(t)$ it also holds that $V(x,t) < \Omega(x,t)$, which contradicts $x \geq x^{\star}(t)$ being the continuation region. A similar argument applies for the other inequality.

But this condition is not sufficient to solve the problem, since the value of $x^{\star}(t)$ is still unknown. The condition that allows us to solve the problem (of jointly finding V and $x^{\star}$) is to impose further smoothness to our value function, it must not only be continuous, but continuously differentiable. This condition is called "smooth pasting" and it requires the first derivative of the value function to be continuous, that is:

$$V_x(x^{\star}(t), t) = \Omega_x(x^{\star}(t), t)$$

The reason behind the smooth pasting condition is not at all evident. I will illustrate it below, but I recommend checking Appendix C of Chapter 4 in Dixit and Pindyck (1994), or Stokey (2009, Prop 6.4 pg 124).

To see why the smooth pasting condition arise consider the following case built for a contradiction: the value matching condition holds, but the smooth pasting condition fails, hence V and Ω must meet at a kink. There are two options:

i. There is an upward kink (forming a concave function). If this is the case then, by continuity, Ω would be higher than F for some value $x > x^{\star}$. Contradicting that the continuation region starts at $x^{\star}$.

ii. There is a downward kink (forming a convex function). If this is the case then $x^{\star}$ cannot be a point of indifference either. There is a better strategy, namely continuing for some time Δt and then choosing what to do. This strategy give higher (expected) payoff.
To see this recall the random walk formulation of the brownian motion, in a time lapse Δt x can either go up by h with probability p or down by $-h$ with probability $1-p$, where:
$$h = \sigma\sqrt{\Delta t} \qquad p = \frac{1}{2}\left(1 + \frac{\mu}{\sigma}\sqrt{\Delta t}\right)$$
Then the agent can continue if the step is upward and stop if it is downward. The expected payoff of this strategy is:
$$V(x^{\star}(t), t) = \pi(x^{\star}(t), t)\,\Delta t + \frac{1}{1+\rho\Delta t}\left[pV(x^{\star}(t)+h, t+\Delta t) + (1-p)\,\Omega(x^{\star}(t)-h, t+\Delta t)\right]$$

We can take a Taylor expansion around $(x^\star(t), t)$ to approximate the value of $V(x^\star(t)+h, t+\Delta t)$ and $\Omega(x^\star(t)-h, t+\Delta t)$:

$$\begin{aligned} V(x^\star(t)+h, t+\Delta t) &\approx V(x^\star(t), t) + V_x(x^\star(t), t)\, h + V_t(x^\star(t), t)\, \Delta t \\ \Omega(x^\star(t)-h, t+\Delta t) &\approx \Omega(x^\star(t), t) - \Omega_x(x^\star(t), t)\, h + \Omega_t(x^\star(t), t)\, \Delta t \end{aligned}$$

Replacing gives:

$$V(x^\star(t), t) = \pi(x^\star(t), t)\, \Delta t + \frac{1}{1+\rho\Delta t}\left(V(x^\star(t), t) + \frac{1}{2}\left[\left(V_x(x^\star(t), t) - \Omega_x(x^\star(t), t)\right) h + (V_t(x^\star(\right.\right.$$

where we use the value matching condition and the fact that $ph \approx \frac{1}{2}\sigma\sqrt{\Delta t}$ and $p\Delta t \approx \frac{1}{2}\Delta t$.

Note that what matters for evaluating the strategy is the continuation value, and that Δt is of order h^2, so the first two terms in the continuation value $\left(V(x^\star(t), t) + \frac{1}{2}h\left(V_x(x^\star(t), t) - \Omega_x(\right.\right.$ will dictate the behavior of the gain as $\Delta t \to 0$ (or equivalently $h \to 0$). These terms are positive as long as $V_x(x^\star(t), t) > \Omega_x(x^\star(t), t)$, which is the case if there is a downward kink.

Then there cannot be a downward kink, since it would contradict the optimality of the strategy of stopping at $x^\star(t)$.

Example 7.3. Consider a firm that has flow revenues of e^{x_t}, and that can be closed at any time and sold for a value $\Omega > 0$. The owner of the firm is risk neutral and discounts the future at a rate $\rho > 0$. x_t follows:

$$dx_t = \mu dt + \sigma dW$$

The problem of the firm's owner is then:

$$V(x) = \max\left\{\Omega, e^x \Delta t + \frac{1}{1+\rho\Delta t} E\left[V(x+dx)\right]\right\}$$

where continuation is optimal for $x \geq x^\star$. Note that the problem is independent of time.

As long as x is in the continuation region the value function satisfies the HJB equation:

$$\rho V(x) = e^x + \mu V_x(x) + \frac{1}{2}\sigma^2 V_{xx}(x)$$

This is a second order ordinary differential equation with constant coefficients. Then we know that the solution has the form:

$$V(x) = V^P(x) + A_1 H_1(x) + A_2 H_2(x)$$

where V^P is a particular solution to the differential equation, H_1 and H_2 are homogenous solutions, and $A_1$and A_2 are constants to be determined.

The particular solution is easy to obtain. We can solve for the value of never stopping:

$$V^P(x) = E\left[\int_0^\infty e^{-\rho t} e^x dt\right]$$

We can solve this expectation using the results in example 6.1. We get:

$$V^P(x) = \frac{x}{\rho - \left(\mu + \frac{1}{2}\sigma^2\right)}$$

we assume that $\rho - \left(\mu + \frac{1}{2}\sigma^2\right) > 0$ in order to guarantee the existence of this solution.

The homogenous solutions are obtained from the homogenous equation:

$$\rho H(x) = \mu H_x(x) + \frac{1}{2}\sigma^2 H_{xx}(x)$$

by guessing that $H(x) = e^{\xi x}$ and replacing we get:

$$\begin{aligned} \rho e^{\xi x} &= \mu \xi e^{\xi x} + \frac{1}{2}\sigma^2 \xi^2 e^{\xi x} \\ 0 &= -\rho + \mu \xi + \frac{1}{2}\sigma^2 \xi^2 \end{aligned}$$

our guess is verified for ξ a root of the equation above. There are two roots:

$$\xi_2 = -\frac{\mu + \sqrt{\mu^2 + 2\sigma^2 \rho}}{\sigma^2} \qquad \xi_2 = \frac{-\mu + \sqrt{\mu^2 + 2\sigma^2 \rho}}{\sigma^2}$$

note that $\xi_1 < 0 < 1 < \xi_2$, this follows from $\rho > 0$ and the assumption $\rho - \left(\mu + \frac{1}{2}\sigma^2\right) > 0$. Joining we get the solution for our HJB equation:

$$V(x) = V^P(x) + A_1 e^{\xi_1 x} + A_2 e^{\xi_2 x}$$

Now we must determine the values of A_1 and A_2. To do so we first need to impose certain conditions on our value function.

i. From optimality in exit it must be that: $V(x) \geq \Omega$.

ii. From feasibility it must be that: $V(x) \leq V^P(x) + \Omega$.

iii. Value matching implies: $V(x^\star) = \Omega$.

We will show that $A_2 = 0$. Suppose for a contradiction that $A_2 > 0$, then as $x \to \infty$ we have $e^{\xi_1 x} \to 0$ (since $\xi_1 < 0$), and $e^{\xi_2 x} \to \infty$ (since $\xi_2 > 0$), since $A_2 > 0$ this implies that V violates its upper bound. Now suppose for a contradiction that $A_2 < 0$, as before $e^{\xi_1 x} \to 0$ and $e^{\xi_2 x} \to \infty$, since $\xi_2 > 1$ the last term will grow faster than the first one, thus violating the lower bound (the value goes to $-\infty$). Then it must be that $A_2 = 0$.

Then we can obtain A_1 from the value matching condition:

$$\begin{aligned} V(x^\star) &= V^P(x^\star) + A_1 e^{\xi_1 x^\star} \\ \left(\Omega - V^P(x^\star)\right) e^{-\xi_1 x^\star} &= A_1 \end{aligned}$$

with this the solution is complete, given a value for $x^{\star}$. It is left to find such value, for that we make use of the smooth pasting condition:

$$
\begin{aligned}
V_x\left(x^{\star}\right) &= 0 \\
V_x^P\left(x^{\star}\right) + A_1 \xi_1 e^{\xi_1 x^{\star}} &= 0 \\
\frac{1}{\rho - \left(\mu + \frac{1}{2}\sigma^2\right)} + \left(\Omega - \frac{x^{\star}}{\rho - \left(\mu + \frac{1}{2}\sigma^2\right)}\right)\xi_1 &= 0 \\
\frac{1}{\xi_1} + \left(\rho - \left(\mu + \frac{1}{2}\sigma^2\right)\right)\Omega &= x^{\star}
\end{aligned}
$$

8 The Kolmogorov Forward Equation

The last section of this part of the course develops the Kolmogorov Forward Equation, which describes the dynamics of the probability distribution of a random variable (given its initial value). Moreover, it characterizes the stationary distribution of the variable if such distribution exists. This is of particular importance for models with heterogenous agents since the distribution of the agents in the economy is obtained via the KFE.

Given some initial conditions x_0 and t_0 the objective is to characterize the probability distribution function $\varphi(x,t)$:

$$\Pr(x_t \in [a,b]) = \int_a^b \varphi(u,t)\,du$$

In order to characterize φ we first need to impose a process for x, and then use the random walk approximation. For simplicity:

$$dx = \mu dt + \sigma dW$$

In the random walk approximation the process varies in a period of length Δt by a magnitude of h, it increases with probability p or decreases with probability $1-p$, where:

$$h = \sigma\sqrt{\Delta t} \qquad p = \frac{1}{2}\left(1 + \frac{\mu}{\sigma}\sqrt{\Delta t}\right)$$

From time $t-\Delta t$ to time t the process can reach a value x either by growing from $x-h$ or by decreasing from $x+h$. Then the probability (or more intuitively the fraction of the mass) at point x at time t is given by:

$$\varphi(x,t) = p\varphi(x-h,t-\Delta t) + (1-p)\varphi(x+h,t-\Delta t)$$

We can approximate the elements of the right hand side with a second order Taylor expansion:

$$\varphi(x \pm h, t-\Delta t) \approx \varphi(x,t) - \Delta t\frac{\partial\varphi(x,t)}{\partial t} \pm h\frac{\partial\varphi(x,t)}{\partial x} + \frac{1}{2}h^2\frac{\partial^2\varphi(x,t)}{\partial x^2}$$

Note that terms of order higher than Δt are ignored. We can replace to get:

$$\begin{aligned}
0 &= -\Delta t\frac{\partial\varphi(x,t)}{\partial t} + (1-2p)\left(h\frac{\partial\varphi(x,t)}{\partial x}\right) + \frac{1}{2}h^2\frac{\partial^2\varphi(x,t)}{\partial x^2} \\
0 &= -\Delta t\frac{\partial\varphi(x,t)}{\partial t} - \frac{\mu}{\sigma}\sqrt{\Delta t}\left(\sigma\sqrt{\Delta t}\frac{\partial\varphi(x,t)}{\partial x}\right) + \frac{1}{2}\sigma^2\Delta t\frac{\partial^2\varphi(x,t)}{\partial x^2} \\
0 &= -\frac{\partial\varphi(x,t)}{\partial t} - \mu\frac{\partial\varphi(x,t)}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2\varphi(x,t)}{\partial x^2}
\end{aligned}$$

which gives the KFE:

$$\frac{\partial\varphi(x,t)}{\partial t} = -\mu\frac{\partial\varphi(x,t)}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2\varphi(x,t)}{\partial x^2}$$

If x follows a more general diffusion process we can change the argument above to get:

$$\frac{\partial\varphi\left(x,t\right)}{\partial t}=-\frac{\partial\left[\mu\left(x,t\right)\varphi\left(x,t\right)\right]}{\partial x}+\frac{1}{2}\frac{\partial^{2}\left[\sigma\left(x,t\right)^{2}\varphi\left(x,t\right)\right]}{\partial x^{2}}$$

The KFE is specially useful for finding the stationary distribution of the process. In this case the distribution does not depend on time so the KFE is:

$$0=-\frac{\partial\left[\mu\left(x,t\right)\varphi\left(x\right)\right]}{\partial x}+\frac{1}{2}\frac{\partial^{2}\left[\sigma\left(x,t\right)^{2}\varphi\left(x\right)\right]}{\partial x^{2}}$$

This equation can be integrated once to get:

$$c_{1}=-2\mu\left(x,t\right)\varphi\left(x\right)+\frac{\partial\left[\sigma\left(x,t\right)^{2}\varphi\left(x\right)\right]}{\partial x}$$

where c_1 is a constant of integration (to be determined later). Then we can use the integrating factor:

$$s\left(x\right)=e^{-\int^{x}\frac{2\mu(z,t)}{\sigma^{2}(z,t)}dz}$$

By multiplying both sides by the integrating factor we get:

$$s\left(x\right)c_{1}=e^{-\int^{x}\frac{2\mu(z,t)}{\sigma^{2}(z,t)}dz}\left(-2\mu\left(x,t\right)\varphi\left(x\right)+\frac{\partial\left[\sigma\left(x,t\right)^{2}\varphi\left(x\right)\right]}{\partial x}\right)$$

The RHS can be rewritten noting that:

$$\begin{aligned}\frac{d}{dx}\left[s\left(x\right)\sigma^{2}\left(x,t\right)\varphi\left(x\right)\right]&=\frac{d}{dx}\left[e^{-\int^{x}\frac{2\mu(z,t)}{\sigma^{2}(z,t)}dz}\sigma^{2}\left(x,t\right)\varphi\left(x\right)\right]\\&=\frac{d}{dx}\left[e^{-\int^{x}\frac{2\mu(z,t)}{\sigma^{2}(z,t)}dz}\right]\sigma^{2}\left(x,t\right)\varphi\left(x\right)+s\left(x\right)\frac{d}{dx}\left[\sigma^{2}\left(x,t\right)\varphi\left(x\right)\right]\\&=-\frac{2\mu\left(x,t\right)}{\sigma^{2}\left(x,t\right)}\sigma^{2}\left(x,t\right)\varphi\left(x\right)+s\left(x\right)\frac{d}{dx}\left[\sigma^{2}\left(x,t\right)\varphi\left(x\right)\right]\\&=-2\mu\left(x,t\right)\varphi\left(x\right)+s\left(x\right)\frac{d}{dx}\left[\sigma^{2}\left(x,t\right)\varphi\left(x\right)\right]\end{aligned}$$

Then we get:

$$s\left(x\right)c_{1}=\frac{d}{dx}\left[s\left(x\right)\sigma^{2}\left(x,t\right)\varphi\left(x\right)\right]$$

Integrating again:

$$c_{1}\int^{x}s\left(y\right)dy+c_{2}=s\left(x\right)\sigma^{2}\left(x,t\right)\varphi\left(x\right)$$

rearranging gives:

$$\varphi\left(x\right)=\frac{1}{s\left(x\right)\sigma^{2}\left(x,t\right)}\left(c_{1}\int^{x}s\left(y\right)dy+c_{2}\right)$$

where $\int^{x}f\left(\xi\right)d\xi=F\left(x\right)$, being F the antiderivative of f.

Example 8.1. Dynamics and Barriers Consider a brownian motion with two reflecting barriers $\overline{x}$ and $\underline{x}$. The process behaves as $dx = \mu dt + \sigma dW$ for $x \in (\underline{x}, \overline{x})$, but is kept in those bounds by force. In terms of the random walk representation that means that starting at $\overline{x} - h$ the process stays at $\overline{x} - h$ with probability p, instead of taking a step up, and goes down to $\overline{x} - 2h$ with probability $1 - p$. Similarly for $\overline{x} + h$.

The KFE applies for any point in the interior of the domain, so for $x \in (\underline{x}, \overline{x})$ we have:

$$\frac{\partial \varphi(x,t)}{\partial t} = -\mu \frac{\partial \varphi(x,t)}{\partial x} + \frac{1}{2}\sigma^2 \frac{\partial^2 \varphi(x,t)}{\partial x^2}$$

Moreover, since we are interested in the stationary behavior of the process we know that the distribution does not depend on time, which results in:

$$0 = -\mu \frac{\partial \varphi(x)}{\partial x} + \frac{1}{2}\sigma^2 \frac{\partial^2 \varphi(x)}{\partial x^2}$$

or better:

$$\varphi^{'}(x) = \frac{1}{2}\frac{\sigma^2}{\mu}\varphi^{''}(x)$$

We can solve this equation:

$$\varphi(x) = Ae^{\gamma x} + B$$

where $\gamma = \frac{2\mu}{\sigma^2}$ and A and B are constants to be determined. To find them we can make use of the boundary conditions implied by the barriers.

From the random walk approximation we can derive the following equation for the upper bound:

$$\begin{aligned}\varphi(\overline{x} - h) &= p\varphi(x - h) + p\varphi(x - 2h)\\ (1-p)\varphi(\overline{x} - h) &= p\varphi(x - 2h)\end{aligned}$$

Using now a second order Taylor expansion around $\overline{x} - h$:

$$\begin{aligned}(1-p)\varphi(\overline{x} - h) &= p\left(\varphi(\overline{x} - h) - h\varphi^{'}(\overline{x} - h) + \frac{1}{2}h^2\varphi^{''}(\overline{x} - h)\right)\\ (1-2p)\varphi(\overline{x} - h) &= -ph\varphi^{'}(\overline{x} - h) + p\frac{1}{2}h^2\varphi^{''}(\overline{x} - h)\\ -\frac{\mu}{\sigma}\sqrt{\Delta t}\varphi(\overline{x} - h) &= -\frac{1}{2}\left(1 + \frac{\mu}{\sigma}\sqrt{\Delta t}\right)\sigma\sqrt{\Delta t}\varphi^{'}(\overline{x} - h) + \frac{1}{4}\left(1 + \frac{\mu}{\sigma}\sqrt{\Delta t}\right)\sigma^2\Delta t\varphi^{''}(\overline{x} - h)\\ -\frac{2\mu}{\sigma^2}\varphi(\overline{x} - h) &= -\left(1 + \frac{\mu}{\sigma}\sqrt{\Delta t}\right)\varphi^{'}(\overline{x} - h) + \frac{1}{2}\left(1 + \frac{\mu}{\sigma}\sqrt{\Delta t}\right)\sigma\sqrt{\Delta t}\varphi^{''}(\overline{x} - h)\end{aligned}$$

taking $\Delta t \to 0$ we get:

$$\begin{aligned}\frac{2\mu}{\sigma^2}\varphi(\overline{x}) &= \varphi^{'}(\overline{x})\\ \gamma\varphi(\overline{x}) &= \varphi^{'}(\overline{x})\end{aligned}$$

Replacing for the solution of φ we find that $B = 0$. Then A is found to guarantee that φ integrates to one. This results in:

$$\varphi(\overline{x}) = \frac{\gamma e^{\gamma x}}{e^{\gamma\overline{x}} - e^{\gamma\underline{x}}}$$

Part III
Applications

Explain here this part of the course

9 Real Options

Consider the problem of a firm that is thinking about investing in a new project. The payoff that the project generates varies stochastically, but its cost is fixed. To be precise: the firm can, at any point in time, pay a fixed cost I to invest on a project that will have a payoff $x(t)$. Firm's investment opportunity is a perpetual call option, that is, the right but not the obligation to buy a share of some asset at a pre-specified price.

The payoff is assumed to follow a geometric brownian motion, so that:

$$dx = \mu x dt + \sigma x dW$$

The firm discounts the future at a rate ρ, so the problem of the firm is:

$$V(x_0) = \max_T E\left[(x(T) - I)e^{-\rho T} | x(0) = x_0\right]$$

To fix ideas we can first solve for the deterministic case. For this we set $\sigma = 0$, which implies that $x(T) = x_0 e^{\mu t}$, for some initial value x_0. Then:

$$V(x_0) = \max_T \left(x_0 e^{\mu T} - I\right) e^{-\rho T}$$

The following results follow:

i. If $\mu \leq 0$ then the payoff x is decreasing (or constant), so it is better to invest immediately if $x_0 > I$, or never to invest if $x_0 \leq I$. This implies that:

$$V(x_0) = \max\{x_0 - I, 0\}$$

ii. If $0 < \mu \leq \rho$ then x is growing, so the value of the firm (the value of holding the option to invest) is positive, even if initially $x_0 < I$. Eventually $x > I$.

(a) The optimal time is given by:

$$\begin{aligned}
\frac{\partial\left(x_0 e^{\mu T} - I\right)e^{-\rho T}}{\partial T} &= 0 \\
-(\rho - \mu)x_0 e^{-(\rho-\mu)T} + \rho I e^{-\rho T} &= 0 \\
\frac{1}{\mu}\ln\frac{\rho I}{(\rho-\mu)x_0} &= T \\
T = \max\left\{\frac{1}{\mu}\ln\frac{\rho I}{(\rho-\mu)x_0}, 0\right\}
\end{aligned}$$

(b) In some cases it is best to invest immediately. This happens if:

$$\begin{aligned}
\frac{1}{\mu}\ln\frac{\rho I}{(\rho-\mu)x_0} &\leq 0 \\
\frac{\rho I}{(\rho-\mu)x_0} &\leq 1 \\
x^\star = \frac{\rho I}{\rho-\mu} &\leq x_0
\end{aligned}$$

Note that higher μ increases the threshold value of x_0. Thus inducing longer waits.

(c) Joining we get:

$$V\left(x_0\right)=\begin{cases}\frac{\mu}{\rho-\mu}I\left(\frac{(\rho-\mu)x_0}{\rho I}\right)^{\frac{\rho}{\mu}} & \text{if } x_0\leq\frac{\rho I}{\rho-\mu}\\ x_0-I & \text{otw}\end{cases}$$

iii. If $\mu>\rho$ then the payoff grows faster than the firm discount of the future, which implies that the firm wants to wait forever.

Now we can solve the stochastic version of the problem. It is no longer possible to find $T^{\star}$ directly, but we can still find the threshold value $x^{\star}$. To do it we first define the HJB equation, recall from equation (7.9) that:

$$\rho V dt=E\left[dV\right]$$

(noting that the instantaneous payoff before investing is zero). We can use Ito's lemma to expand the RHS:

$$\rho V=\mu x V'+\frac{1}{2}\sigma^2x^2V''$$

There are three boundary conditions that must hold:

$$V\left(0\right)=0\qquad V\left(x^{\star}\right)=x^{\star}-I\qquad V'\left(x^{\star}\right)=1$$

The first one follows from 0 being an absorbing state (because of the properties of the geometric brownian motion). The second one is value matching and the third one is smooth pasting.

We guess that the solution is of the form:

$$V\left(x\right)=Ax^{\beta}$$

for some A and β to be found later. This clearly solves the HJB equation. Replacing we can solve for β:

$$\begin{aligned}\rho V&=\mu x V'+\frac{1}{2}\sigma^2x^2V''\\ \rho A x^{\beta}&=\mu\beta A x^{\beta}+\frac{1}{2}\sigma^2A\beta\left(\beta-1\right)x^{\beta}\\ \rho&=\mu\beta+\frac{1}{2}\sigma^2\beta\left(\beta-1\right)\\ 0&=-\rho+\left(\mu-\frac{1}{2}\sigma^2\right)\beta+\frac{1}{2}\sigma^2\beta^2\end{aligned}$$

β is then found from the roots of this equation:

$$\beta_1=\frac{1}{2}-\frac{\mu}{\sigma^2}+\sqrt{\left(\frac{1}{2}-\frac{\mu}{\sigma^2}\right)^2+2\frac{\rho}{\sigma^2}}>1$$

$$\beta_2=\frac{1}{2}-\frac{\mu}{\sigma^2}-\sqrt{\left(\frac{1}{2}-\frac{\mu}{\sigma^2}\right)^2+2\frac{\rho}{\sigma^2}}<0$$

Since there are two distinct roots the solution to the HJB equation is in general:

$$V(x) = A_1 x^{\beta_1} + A_2 x^{\beta_2}$$

But in order for the first boundary condition to hold we need that $A_2 = 0$, since, with $\beta_2 < 0$, we could not evaluate the function otherwise. This leaves us with only one root, which we denote β, and one constant A that we find below.

Replacing on the value matching and smooth pasting conditions we get:

$$A(x^\star)^\beta = x^\star - I \qquad \beta A (x^\star)^{\beta-1} = 1$$

Solving for $A = \frac{(x^\star)^{1-\beta}}{\beta}$ and replacing on the first equation we get:

$$\begin{aligned} x^\star &= \beta x^\star - \beta I \\ x^\star &= \frac{\beta I}{\beta - 1} \end{aligned}$$

which also gives the value of A.

Note that in the optimal strategy the firm does not invest when $x^\star$ is equal to I (when the net present value of investing becomes positive), but instead there is wedge between the cost of investing and the value of investing. The wedge is given because the firm has to be compensated for giving up the option to wait and see if the value increases even further.

Comparative Statics [Optional]

The threshold value $x^\star$ depends on the parameters of the model through β. Although we have an explicit solution for β in this case, that is not always the case. Nevertheless we can use the quadratic equation that gives rise to β to run comparative statics.

Let Q be the quadratic equation, so that:

$$Q = -\rho + \left(\mu - \frac{1}{2}\sigma^2\right)\beta + \frac{1}{2}\sigma^2\beta^2$$

we want to know how β depends on σ. Taking total differentials we get:

$$\frac{dQ}{d\sigma} = \frac{\partial Q}{\partial \beta}\frac{d\beta}{d\sigma} + \frac{\partial Q}{\partial \sigma} = 0$$

where the derivatives are evaluated at the positive root β found above. This expression gives:

$$\frac{d\beta}{d\sigma} = -\frac{\left(\frac{\partial Q}{\partial \sigma}\right)}{\left(\frac{\partial Q}{\partial \beta}\right)}$$

Signing the numerator is easy, since $\frac{\partial Q}{\partial \sigma} = \sigma\beta(\beta - 1) > 0$, we know it is positive since the positive root β is higher than 1. Signing the denominator requires us to know the shape of Q. It can be easily shown that Q is increasing at β_1:

$$\frac{\partial Q}{\partial \beta} = \mu - \frac{1}{2}\sigma^2 + \sigma^2\beta = \sigma^2\sqrt{\left(\frac{1}{2} - \frac{\mu}{\sigma^2}\right)^2 + 2\frac{\rho}{\sigma^2}} > 0$$

Then:

$$\frac{d\beta}{d\sigma} < 0$$

This means that higher variance (more uncertainty over the payoff of investing) reduces β, which in turn increases $\frac{\beta}{\beta-1}$. So the wedge between $x^{\star}$ and I increases with uncertainty, in other words the firm will need a larger return on the investment in order to invest.

10 Menu cost Stokey (2009, Ch. 7)

Consider a firm whose profit flow at any date t depends on its relative price, that is: the ratio of its own nominal price to an aggregate (industry-wide or economy-wide) price index, where the latter is a geometric Brownian motion. Recall that if the price follows a GBM then its log follows a brownian motion. It is then convenient to work with the prices in log form. Let $p(t)$ be the log of the firm's nominal price and $\overline{p}(t)$ the log of the aggregate price index. Then:

$$d\overline{p} = -\mu dt + \sigma dW_p$$

The initial value for the firm's (log) nominal price p_0 is given. The firm can change its nominal price at any time, but to do so it must pay a fixed adjustment cost $c > 0$. This cost is constant over time and measured in real terms. Because control entails a fixed cost, the firm adjusts the price only occasionally and by discrete amounts.

The problem of the firm is to choose when to adjust the price, and by how much. One can see this as a problem of choosing the (random) times at which to adjust the price, or of choosing an inaction region, such that the price is adjusted when some condition is met.

Since the profit flow at any date depends only on the firm's relative price, the problem can be formulated in terms of that one state variable. Let:

$$z(t) = p(t) - \overline{p}(t)$$

When the firm adjusts its price the variable z jumps. Part of the problem will be to find the optimal value $z^\star$ to which z is set when the firm decides to take action. Between adjustments z evolves only with $\overline{p}$, so we have (for any time at which there is no adjustment):

$$dz = \mu dt + \sigma dW$$

where $dW = -dW_p$.

The profit flow of the firm, $\pi(z)$, is a stationary function of its relative price z, and profits are discounted at a constant interest rate r. The following restrictions on π, r, c and the parameters μ, σ^2 insure that the problem is well behaved:

i. $r, c, \sigma^2 > 0$

ii. π is continuous everywhere, strictly increasing on $(-\infty, 0)$ and strictly decreasing on $(0, \infty)$.

   (a) The location of the peak of π at 0 is arbitrary.

We will assume that π takes the following form:

$$\pi(z) = \begin{cases} \pi_0 e^{\eta_+ z} & \text{if } z \geq 0 \\ \pi_0 e^{\eta_- z} & \text{if } z < 0 \end{cases}$$

where $\eta_+ < 0 < \eta_-$.

These assumptions imply that it is optimal to change the price if z gets too low or too high. Then the inaction region is: $(\underline{z}, \bar{z})$.

The HJB equation for $z \in (\underline{z}, \bar{z})$ is:

$$\rho V(z) = \pi(z) + \mu V'(z) + \frac{1}{2}\sigma^2 V''(z)$$

The boundary conditions for V are value matching and smooth pasting at $\underline{z}$ and $\overline{z}$:

$$\begin{aligned} V(\underline{z}) &= V(z^\star) - c \\ V(\overline{z}) &= V(z^\star) - c \\ V'(\underline{z}) &= 0 \\ V'(\bar{z}) &= 0 \end{aligned}$$

while $z^\star$ is optimally found to maximize V. So it must satisfy:

$$V'(z^\star) = 0$$

The solution to the HJB equation is, just as before:

$$V(z) = V^p(z) + A_1 e^{\xi_1 z} + A_2 e^{\xi_2 z}$$

where V^p is a particular solution and $H(z) = e^{\xi z}$ is a solution to the homogeneous equation.

Finding the particular solution is not trivial. Stokey proposes the following solution:

$$W(z) = E\left[\int_0^T e^{-rt}\pi(z(t))\,dt\right]$$

where T is the (random) time at which the price will be adjusted. W gives then the expected discounted value of the profits until the next adjustment. This function is difficult to deal with, since T is not a real number, but instead a random variable. It is however possible to exploit this to express W as an integral over values of z. This goes beyond what we are covering, the details are available in Stokey (2009, Sec. 3.5). Critically it can be shown that:

$$W(\underline{z}) = W(\bar{z}) = 0$$

since there is no time until the next adjustment. It will occur in that instant.

This simplifies the value matching conditions to:

$$\begin{aligned} A_1 e^{\xi_1 \underline{z}} + A_2 e^{\xi_2 \underline{z}} &= V(z^\star) - c \\ A_1 e^{\xi_1 \overline{z}} + A_2 e^{\xi_2 \overline{z}} &= V(z^\star) - c \end{aligned}$$

Unfortunately we cannot further solve this problem.

References

Abreu, D., Milgrom, P., and Pearce, D. (1991). Information and timing in repeated partnerships. *Econometrica*, 59(6):1713–1733.

Bonatti, A. and Horner, J. (2011). Collaborating. *American Economic Review*, 101(2):632–63.

Dixit, A. (1993). *The Art of Smooth Pasting*. Fundamentals of pure and applied economics. Harwood Academic Publishers.

Dixit, A. K. and Pindyck, R. S. (1994). *Investment under Uncertainty*. Princeton University Press.

Ely, J. C. (2017). Beeps. *American Economic Review*, 107(1):31–53.

Fudenberg, D. and Tirole, J. (1991). *Game Theory*. MIT Press.

Georgiadis, G. (2015). Projects and team dynamics. *The Review of Economic Studies*, 82(1):187.

Holmstrom, B. (1999). Managerial incentive problems: A dynamic perspective. *The Review of Economic Studies*, 66(1):169.

Holmstrom, B. and Milgrom, P. (1987). Aggregation and Linearity in the Provision of Intertemporal Incentives. *Econometrica*, 55(2):303–328.

Horner, J. and Skrzypacz, A. (2016). Selling information. *Journal of Political Economy*, 124(6):1515–1562.

Irigoyen, C., Rossi-Hansberg, E., and Wright, M. (2009). *Solutions Manual for 'Recursive Methods in Economic Dynamics'*. Harvard University Press.

Kamenica, E. and Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, 101(6):2590–2615.

Klette, T. and Kortum, S. (2004). Innovating firms and aggregate innovation. *Journal of Political Economy*, 112(5):986–1018.

Kolb, A. (2017). Strategic real options. Job Market Paper.

Kolmogorov, A. and Fomin, S. (1999). *Elements of the Theory of Functions and Functional Analysis*. Number v. 1 in Dover books on mathematics. Dover.

Kolmogorov, A. and Fomin, S. (2012). *Introductory Real Analysis*. Dover Books on Mathematics. Dover Publications.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, 53(6):1315–1335.

Lucas, R. E. (1978). On the size distribution of business firms. *The Bell Journal of Economics*, 9(2):508–523.

Luttmer, E. G. (2012). Technology diffusion and growth. *Journal of Economic Theory*, 147(2):602 – 622. Issue in honor of David Cass.

Luttmer, E. G. J. (2007). Selection, growth, and the size distribution of firms*. *The Quarterly Journal of Economics*, 122(3):1103.

Luttmer, E. G. J. (2011). On the mechanics of firm growth. *The Review of Economic Studies*, 78(3):1042.

Mas-Colell, A., Whinston, M. D., and Green, J. R. (1995). *Microeconomic Theory*. Oxford University Press.

Milgrom, P. and Segal, I. (2002). Envelope Theorems for Arbitrary Choice Sets. *Econometrica*, 70(2):583–601.

Moll, B. (2014). Productivity losses from financial frictions: Can self-financing undo capital misallocation? *American Economic Review*, 104(10):3186–3221.

Morris, S. and Shin, H. S. (2003). Global ggame: Theory and applications. In Dewatripont, M., Hansen, L., Turnovsky, S., and Congress, E. S. W., editors, *Advances in Economics and Econometrics: Theory and Applications : Eighth World Congress*, Econometric Society monographs. Cambridge University Press.

Øksendal, B. (2003). *Stochastic differential equations*. Springer.

Rahman, D. (2012). But who will monitor the monitor? *American Economic Review*, 102(6):2767–97.

Rockafellar, R. (1997). *Convex Analysis*. Convex Analysis. Princeton University Press.

Sannikov, Y. and Skrzypacz, A. (2007). Impossibility of collusion under imperfect monitoring with flexible production. *American Economic Review*, 97(5):1794–1823.

Stokey, N., Lucas, R., and Prescott, E. (1989). *Recursive Methods in Economic Dynamics*. Harvard University Press.

Stokey, N. L. (2009). *The Economics of Inaction: Stochastic Control Models with Fixed Costs*. Princeton University Press.

Sundaram, R. (1996). *A First Course in Optimization Theory*. Cambridge University Press.

Topkis, D. (1998). *Supermodularity and Complementarity*. Frontiers of Economic Research. Princeton University Press.

Wade, W. (2010). *An Introduction to Analysis*. Featured Titles for Real Analysis Series. Prentice Hall/Pearson.