UTRECHT UNIVERSITY

BACHELOR THESIS

FACULTY OF BETA SCIENCES

Entropy and Option Pricing

Author
M.N. MOOIJ
5746604

Supervisor
Dr. K. DAJANI

14th June, 2019

Abstract

This thesis will look at entropy and its applications to option pricing. We will discuss ergodic theory and its connection with entropy. In the second part of this thesis we use entropy to derive an option pricing model. In the end we will use Python to compare this model with the Black-Scholes model. We see that the entropy model gives the same results as the Black-Scholes model.

Contents

Introduction

In physics, the concept of entropy has been introduced in 1865 by Rudolf Clausius to give meaning to the amount of disorder in a thermodynamical system. For over 80 years this physical definition was the only one. Until in 1948 Claude Shannon introduced the notion of Information entropy. With this definition he introduced a way to measure chaos in mathematical systems. Ever since entropy has been used in numerous fields, in particular the field of information theory.

This thesis discusses an economical application of entropy. We will use the notion of entropy to find an alternative model to the Black-Scholes model. This is important scientific work since a different model could result in different prices. This means that we could have two different ways to value options.

In chapter one of this thesis we will give an introduction to measure theory. Then we will use this to define ergodic theory. In the second part of this chapter, we will prove some results in ergodic theory. In the end of this chapter we will prove Birkhoff's Ergodic Theorem. We will end this chapter with an application of this theorem.

In the second chapter we will use ergodic theory to introduce the concept of entropy. We will first do this for an independent transmission source, and then for a general dynamical system. After proving some results regarding entropy, we prove that entropy is isomorphism invariant.

In the third chapter we derive a model for the pricing of European options. We do this by solving a system of equations. In the first half of this chapter we derive a model for a general stock. In the second part of this chapter we derive a model for option pricing, under the assumption that our stock follows a square root process.

In chapter four of this thesis, we will numerically calculate option prices, under the entropy model derived in chapter three. We will compare these prices with the prices calculated under the Black and Scholes model. We will also compare these prices with the real market prices of the options.

1 Ergodic Theory

Ergodic theory has many applications. For example in statistical physics, the mathematics of ergodic theory can be used to measure the speed of a particle.

This chapter gives an introduction to ergodic theory and then derives some important theorems in ergodic theory.

In order to define what ergodic theory is, we will first treat some prerequisites of measure theory. If one is already familiar with measure theory, one can start reading this thesis at section 1.2.

1.1 Measure theory

To get a good basis for ergodic theory, one first needs to study some measure theoretical mathematics. To talk about measure theory we first need to state some basic definitions.

When we think about a measure in $\mathbb{R}^2$, intuitively a measure is then a function which assigns to every shape of $\mathbb{R}^2$ a number which "corresponds" with how "big" this shape is. This raises the question of which shapes can be quantified in this way. This question can be answered with the definition of a σ-algebra.

Definition 1.1.1 (σ - algebra)**.** A σ- algebra $\mathcal{A}$ on a set X is a collection of subsets of X with the following properties:

(i) $X \in \mathcal{A}$

(ii) If $A \in \mathcal{A}$ then also $A^c \in \mathcal{A}$

(iii) Let β be a countable collection of subsets $A \in \mathcal{A}$. Then also $\bigcup\limits_{A \in \beta} A \in \mathcal{A}$

The pair $(X, \mathcal{A})$ is called a measurable space and the elements of $\mathcal{A}$ are called measurable sets.

For example when we look at a set X we can take $A \subset X$ and define our σ-algebra as $\mathcal{A} = \left\{ \emptyset, A, A^c, X \right\}$. One can easily check that this example meets all the requirements of a σ- algebra. So $\mathcal{A}$ gives information about whether an element $x \in X$ belongs to A or not.

When we compare this with $\mathbb{R}^2$ we can take $\mathcal{A} = \mathcal{P}(\mathbb{R}^2)$. Intuitively, this means that with this σ-algebra we can measure every subset of $\mathbb{R}^2$. When we consider the space $\mathbb{R}$ there is an important σ-algebra we need to define.

Definition 1.1.2 (Borel σ-algebra)**.** Let X be a space. The Borel σ-algebra on X, denoted by $B(X)$, is the σ-algebra generated by the closed (or equivalently open) sets on X.

This means that $B(\mathbb{R})$ contains at least all the sets of the form $[a, b]$ such that $a < b$ and that it contains all possible unions of these sets.

Now we are ready to give a definition of a measure.

Definition 1.1.3 (Measure)**.** A measure μ is a function from a σ-algebra $\mathcal{A}$ on X to $[0, \infty]$ with the following properties:

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

(ii) Let β be a countable collection of disjoint subsets $A \in \mathcal{A}$. Then $\mu(\bigcup\limits_{A \in \beta} A) = \sum\limits_{A \in \beta} \mu(A)$

Example 1.1.4. We take as an example a set X equipped with a σ-algebra $\mathcal{A}$ and a measure μ defined by

$$\begin{aligned} \mu : \mathcal{A} &\to \mathbb{R} \\ A &\mapsto |A| \end{aligned}$$

Since the cardinality of a set is always greater or equal to zero we know that the range of μ is indeed $[0, \infty]$.

We can easily check that:

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

(ii) Let β be a countable collection of disjoint subsets $A \in \mathcal{A}$. Then $\mu(\bigcup\limits_{A \in \beta} A) = |\bigcup\limits_{A \in \beta} A| = \sum\limits_{A \in \beta} \mu(A)$ where the last step is true since all $A \in \beta$ are disjoint.

So we see that the function μ is indeed a measure. The name of this particular measure is the counting measure. $\triangle$

Another important measure is the Lebesgue measure. We will only define the Lebesgue measure on $(\mathbb{R}, B(\mathbb{R}))$.

Definition 1.1.5 (Lebesgue measure on $\mathbb{R}$)**.** Let $E \subseteq B(\mathbb{R})$. Let $l : B(\mathbb{R}) \to \mathbb{R}$ be the function that defines the length of an interval in $\mathbb{R}$ i.e. $l([a, b]) = b - a$. We define the Lebesgue measure $\lambda : B(\mathbb{R}) \to [0, \infty]$ as:

$$\lambda(E) = \inf \left\{ \sum_{k=1}^{\infty} l(I_k) \mid (I_k)_{k \in \mathbb{N}} \text{ is a sequence of open intervals such that } E \subseteq \bigcup_{k=1}^{\infty} I_k \right\}$$

We directly see that the Lebesgue measure assigns the value zero to all singletons.

With these definitions we can introduce the notion of a measure space. We call a set X together with a σ-algebra $\mathcal{A}$ on X and a measure μ on X a measure space. We denote this space by $(X, \mathcal{A}, \mu)$.

If $\mu(X) = 1$ we call $(X, \mathcal{A}, \mu)$ a probability space.

Similar to metric and topological spaces, we can also work with functions on measurable spaces. However we need to be more careful since we are working with σ-algebras and want to make sure that we still can measure every subset after we have mapped it with our function. This leads to the following definition.

Definition 1.1.6 (Measurable function)**.** Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces. A function $f : (X, \mathcal{A}) \to (Y, \mathcal{B})$ is said to be measurable if for all $B \in \mathcal{B}$ we have $f^{-1}(B) \in \mathcal{A}$

With this definition we make sure that the functions used preserve the structure of the space.

From now on, when we talk about a function from one measurable space to another measurable space, we assume this function is measurable.

Now we know what is meant with a measurable function, we can progress by defining what we mean by a measure preserving function.

Definition 1.1.7 (Measure preserving function)**.** Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be measure spaces. A function $f : (X, \mathcal{A}, \mu) \to (Y, \mathcal{B}, \nu)$ is measure preserving if for all $B \in \mathcal{B}$ we have $\mu(f^{-1}(B)) = \nu(B)$

Intuitively, what we mean by this is that the "size" of a subset of Y is the same as the size of the preimage of this subset by f.

An important aspect of measure theory is that we sometimes deal with properties that occur for almost all points in the space we are dealing with. In first instance it might be surprising that we talk about this almost-everywhere property but we will make this rigorous now.

Let $(X, \mathcal{A}, \mu)$ be a measure space. We say that the subset $A \in \mathcal{A}$ is negligible if $\mu(A) = 0$. To illustrate this we look at the following example.

Example 1.1.8. Let $(\mathbb{R}, B(\mathbb{R}), \lambda)$ be a measure space with $B(\mathbb{R})$ the Borel σ-algebra on $\mathbb{R}$ and λ the Lebesgue measure on $\mathbb{R}$. We define

$$f : \mathbb{R} \to \mathbb{R}$$

$$f(x) = \begin{cases} 1 \text{ if } x \in \mathbb{Q} \\ 0 \text{ if } x \in \mathbb{R} \backslash \mathbb{Q} \end{cases}$$

Since $\mathbb{Q}$ is a countable collection of singletons we know that $\lambda(\mathbb{Q}) = 0$. This means that the subset $\{x \in X \mid f = 1\}$ is negligible with respect to λ.

△

This gives rise to the definition of μ almost-everywhere occurence. We say a property P occurs μ almost everywhere if the set where P does not occur is negligible. To make this rigorous, let $(X, \mathcal{A}, \mu)$ be a measure space and P some property. Let E be the set on X where P occurs. If E is measurable or a subset of a measurable set, then P occurs μ almost-everywhere if $\mu(E^c) = \mu(x \in X \text{ such that P does not occur}) = 0$.

To end this paragraph we will state the definition of the L^p space. This is the space with p-integrable functions modulo the null sets of the measure we are working with.

Definition 1.1.9. Let $(X, \mathcal{A}, \mu)$ be a measure space. Let $\mathcal{L}^p(X, \mathcal{A}, \mu)$ be the space of p-integrable functions on $(X, \mathcal{A}, \mu)$, with the p-norm defined as usual $\|f\|_p = \left(\int_X |f|^p \, d\mu\right)^{1/p}$. Let $\mathcal{N}(X, \mathcal{A}, \mu)$ be the space of negligible sets on the measure space $(X, \mathcal{A}, \mu)$. Then the space $L^p(X, \mathcal{A}, \mu)$ is defined as $L^p(X, \mathcal{A}, \mu) = \mathcal{L}^p(X, \mathcal{A}, \mu)/\mathcal{N}(X, \mathcal{A}, \mu)$.

1.2 Introduction to Ergodic theory

This section is based on the following notes: *Introduction to Ergodic Theory*[4] and *Ergodic Theory and Entropy*[1].

Ergodic theory is consistently dealing with so called dynamical systems. With the work we have done in the last paragraph the notion of a measure space is easily extended to that of a dynamical system.

Definition 1.2.1 (Dynamical system)**.** A dynamical system is a set X with a σ-algebra $\mathcal{A}$, a probability measure μ and a measurable transformation $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$. We denote this dynamical system by $(X, \mathcal{A}, \mu, T)$. If T is a measure preserving transformation we call $(X, \mathcal{A}, \mu, T)$ a measure preserving dynamical system.

Ergodic theory is about the behaviour of a dynamical system under certain mappings. Since this mapping T is from X to X, the first question that comes to mind is: What is T doing with subsets of X? Or more specifically: Which sets are left invariant by T?

Definition 1.2.2 (T-invariance)**.** Let $(X, \mathcal{A}, \mu, T)$ be a dynamical system. Then $A \in \mathcal{A}$ is said to be T-invariant if $T^{-1}(A)\Delta A$ is a negligible set. With $A\Delta B$ defined as $A\Delta B = (A \cup B)\backslash(A \cap B)$.

What we mean by this in words, is that a set $A \in \mathcal{A}$ is mapped to itself except for $E \subset A$ which are negligible sets. In the remains of this thesis we will see this idea of equality except for null sets come by often.

We are now ready to give a definition of ergodicity. Ergodic measures are merely measures which assign special values to invariant sets. An ergodic measure assigns to invariant sets only the measure zero or the measure one. Intuitively, this means only the entire space and the null sets are left invariant by T.

Definition 1.2.3 (Ergodicity)**.** Let $(X, \mathcal{A}, \mu, T)$ be a dynamical system. We call μ ergodic with respect to T if every T-invariant set has either measure zero or measure one. i.e. If $\forall A \in \mathcal{A}$ such that $\mu(T^{-1}(A)\Delta A) = 0$ we have $\mu(A) = 0$ or $\mu(A) = 1$

Before we show an example of ergodicity we first state and prove a small lemma.

Lemma 1.2.4. *Let $(X, \mathcal{A}, \mu)$ be a probability space and $T : X \to X$ measure preserving. Then the following are equivalent:*

(i) If $f \in L^2(X, \mathcal{A}, \mu)$ with $f(Tx) = f(x)$ for all $x \in X$, then f is a constant almost everywhere.

(ii) T is ergodic

Proof. $(i) \to (ii)$ Suppose $T^{-1}A = A$. We recognize that $\mathbb{1}_A \in L^2(X, \mathcal{A}, \mu)$. Since T is measure preserving we can make the following calculation: $\mathbb{1}_A \cdot T = \mathbb{1}_{T^{-1}A} = \mathbb{1}_A$. By (i) we know this is a constant almost everywhere. So we have two options; $\mu(A) = 1$ or $\mu(A) = 0$. Hence T is ergodic.

$(ii) \to (i)$ Suppose T is ergodic and let $f \in L^2(X, \mathcal{A}, \mu)$ such that $f(Tx) = f(x)$. Let $a \in \mathbb{R}$. Then we know $\{f \geq a\} = \{f \circ T \geq a\} = T^{-1}\{f \geq a\}$. Since T is ergodic we know that the measure of this set is either zero or one.

Now define $a_0 = \sup\{a \in \mathbb{R} \mid \mu(f \geq a) = 1\}$. We can calculate $\mu(f \geq a_0) = \lim_{n\to\infty} \mu(f \geq a_0 - \frac{1}{n}) = \lim_{n\to\infty} 1 = 1$. Now we recognize that $\{f = a_0\} = \{f \geq a_0\} \setminus \cup_{n=1}^{\infty} \{f \geq a_0 + \frac{1}{n}\}$. Since $\mu(f \geq a_0 + \frac{1}{n}) = 0$ for all $n \in \mathbb{N}$ we know that $\mu(f = a_0) = 1$. So f is a constant function. □

Example 1.2.5. With lemma 1.2.4 from above we can work out an illustrating example of ergodicity.
Let $\mathcal{A}$ be the σ-algebra with all Lebesgue measurable sets on $[0, 1)$. (Also called the Lebesgue σ-algebra on $[0, 1)$) and λ the lebesque measure. Let θ be irrational and consider the mapping $T : [0, 1) \to [0, 1)$ defined by $T_\theta x = (x + \theta) mod\ 1$. Now let $f \in L^2([0, 1), \mathcal{A}, \lambda)$ be T-invariant and recognize that we can write the fourier series of this function as:

$$f(x) = \sum_{n\in\mathbb{Z}} a_n e^{2\pi i n x}$$

Since f is T-invariant we also have:

$$f(x) = f(T_\theta x) = \sum_{n\in\mathbb{Z}} a_n e^{2\pi i n(x+\theta)}$$
$$= \sum_{n\in\mathbb{Z}} a_n e^{2\pi i n x} \cdot e^{2\pi i n\theta}$$

If we equal these equations and solve this we get:

$$\sum_{n\in\mathbb{Z}} a_n \left(1 - e^{2\pi i n\theta}\right) e^{2\pi i n x} = 0$$

Since a Fourier expansion has unique coefficients and θ is irrational, we know that $a_n = 0$ for $n \neq 0$. Therefore, $f(x) = a_0$ is a constant. By lemma 1.2.4 we know that T_θ is an ergodic transformation. △

Instead of looking at which mappings preserve the measure of all sets we will now do the exact opposite. We want to look at which measures assign the same "size" to all sets before and after the mapping. This is called an invariant measure.

Definition 1.2.6 (Invariant measure)**.** Let $(X, \mathcal{A}, \mu, T)$ be a dynamical system. We say μ is T-invariant if $\mu(T^{-1}(A)) = \mu(A)$ for all $A \in \mathcal{A}$

To simplify notation in the future we denote by $M(X, \mathcal{A})$ the space of all measures on $(X, \mathcal{A})$ and by $M_{\text{inv}}(X, \mathcal{A}, T) = \left\{\mu \in M \mid \mu \text{ is T-invariant}\right\}$ all the measures that are T-invariant with respect to $(X, \mathcal{A}, T)$. We denote by $M_{\text{erg}}(X, \mathcal{A}, T)$ all the measures that are ergodic with respect to $(X, \mathcal{A}, T)$.
So $M_{\text{erg}}(X, \mathcal{A}, T) = \left\{\mu \in M \mid \text{ T is ergodic w.r.t. } \mu\right\}$.

1.3 Some important results in ergodic theory

In the remains of this chapter we are exploring whether an ergodic measure always exists. It will turn out that with a couple more restrictions on our mapping, an ergodic measure does indeed always exists. To show this, we first state and prove the following lemma, which on first inspection looks as if it has not much to do with ergodic measures, but is crucial to prove our statement. It will turn out that the ergodic measures are precisely the extreme points of M_{inv}.

Definition 1.3.1 (Extreme point)**.** Let X be a topological space and $K \subseteq X$ a convex subset. We call $x \in K$ an extreme point if $K \backslash \{x\}$ is still convex.

Definition 1.3.2 (Face)**.** Let X be a convex vector space. Let K be a nonempty convex subset of X. Then $F \subset K$ is a face of K if for all $x, y \in K$ and $\theta \in (0, 1)$ such that $\theta x + (1 - \theta)y \in F$ then also $x, y \in F$.

The following proof is inspired on the proof in *A course in Functional Analysis*[3], page 142-143.

Lemma 1.3.3 (Krein-Milman Lemma)**.** *Let X be a convex vector space and $K \subset X$ a convex, compact, nonempty subset. Then K has an extreme point.*

Proof. Let $\mathcal{F}$ be the family of all compact faces in K. We know $\mathcal{F} \neq \emptyset$ because K is a face of itself. Now order these elements in $\mathcal{F}$ by nesting them i.e. $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq ... \subseteq K$. Since this chain is bounded by K, we know by Zorn's lemma that it contains a minimal element. We have called this element $\mathcal{F}_0$. Now we want to prove that $\mathcal{F}_0$ is a singleton and a face. This would prove K contains an extreme point.

Assume $\mathcal{F}_0$ contains two points $x, y \in \mathcal{F}_0$. Then by the Hahn-Banach theorem we know that there exists a functional F from $\mathcal{F}_0$ to $\mathbb{R}$ such that $F(x) \neq F(y)$. Since $\mathcal{F}_0 \neq \emptyset$ and $\mathcal{F}_0$ is compact we know by the extreme value theorem that F attains a maximum on K. Let $m = \max\limits_{k \in \mathcal{F}_0} F(k)$.

We define $M = \left\{ k \in \mathcal{F}_0 \mid F(k) = m \right\}$. By the above we know that $M \neq \emptyset$. Also, since $\{m\}$ is closed and F is continuous, we know that $F^{-1}(\{m\}) = M$ is closed. Hence M is compact.
Let $k_1, k_2 \in M$ and $\lambda \in [0,1]$. Let $k = \lambda k_1 + (1-\lambda)k_2$. Since f is linear we know $f(k) = \lambda f(k_1) + (1-\lambda) f(k_2) = m$. So $k \in M$. Therefore, M is convex.

Now let $x \in M$ and suppose $x = \lambda u + (1-\lambda)v$ for some $u, v \in K$ and $\lambda \in [0,1]$. We know that $f(u) \leq m$ and $f(v) \leq m$, so also $m = f(x) = \lambda f(u) + (1-\lambda) f(v) \leq m$. So $\lambda f(u) + (1-\lambda) f(v) = m$. Therefore, we must have $f(u) = f(v) = m$ and $u, v \in M$. So M is an compact face of $\mathcal{F}_0$. Thus this means $M \subsetneq \mathcal{F}_0$(because $y \notin M$). Since we assumed that $\mathcal{F}_0$ was the minimal face of K, this is a contradiction. Therefore, $\mathcal{F}_0$ is a singleton and we know that K contains an extreme point. □

Next, we will prove that the ergodic measures are indeed the extreme points of M_{inv}.

Theorem 1.3.4 (Radon-Nikodym)**.** *Let $(X, \mathcal{A})$ be a measurable space on which two measures $\mu, \nu : \mathcal{A} \to [0, \infty)$ are defined. If $\nu \ll \mu$ then there exists a measurable function $f : X \to [0, \infty)$ such that for any $A \in \mathcal{A}$ we have*

$$\nu(A) = \int_A f \, d\mu \tag{1}$$

This function f is called the Radon-Nikodym derivative and is denoted by $\frac{d\nu}{d\mu}$

A proof of this theorem can be found in *Measure Theory*[2], page 132-135.

Lemma 1.3.5. *Let $(X, \mathcal{A})$ be an arbitrary measurable space and $T : X \to X$ a measurable transformation. If $\mu, \nu \in M_{inv}$ and $\nu \ll \mu$ then the Radon-Nikodym derivative $f = d\nu/d\mu$ is T-invariant in the sense that*

$$f \circ T = f$$

holds μ-almost everywhere.

Proof. Let μ and ν be T-invariant measures. Then by theorem 1.3.4 we know that for any measurable $A \subset X$ we can write

$$\nu(A) = \int_A f \, d\mu$$
$$\nu(A) = \nu(T(A)) = \int_A f \cdot T \, d\mu$$

Therefore, f is T-invariant almost everywhere. □

Theorem 1.3.6. *The extreme points of M_{inv} are exactly the ergodic members of M_{inv}.*

Proof. Instead of proving this, we will prove the contrapositive statement.
Let $\mu \in M_{\text{inv}}$. Assume $\mu \notin M_{\text{erg}}$. Then there exists a $A \in \mathcal{A}$ with $T^{-1}A = A$ and $0 < \mu(A) < 1$. Now let $E \in \mathcal{A}$. We define two measures as $\mu_1(E) = \frac{\mu(E \cap A)}{\mu(A)}$ and $\mu_2(E) = \frac{\mu(E \cap A^c)}{\mu(A^c)}$. Then $\mu_1, \mu_2 \in M_{\text{inv}}$ because A and A^c are both T-invariant and μ is measure preserving. We see that $\mu = \mu(A)\mu_1 + (1 - \mu(A))\mu_2$. So μ is not an extreme point.

Assume $\mu \in M_{\text{erg}}$ is not an extreme point of M_{inv}. Then there exists an $\lambda \in (0, 1)$ such that $\mu = \lambda\mu_1 + (1 - \lambda)\mu_2$ for some $\mu_1, \mu_2 \in M_{\text{inv}}$. We note that when $\mu(A) = 0$ for $A \in \mathcal{A}$ we also have $\mu_1(A) = 0$. Therefore $\mu_1 \ll \mu$. Now we know by Radon-Nikodym (Theorem 1.3.4) that there exists a measurable function $f : X \to \mathbb{R}_{>0}$ such that $\mu_1(A) = \int_A f \, d\mu \; \forall A \in \mathcal{A}$. By Lemma 1.3.5 we know that f is T-invariant μ-almost everywhere.

We want to prove that f is not a constant. If we can prove this we are done, since by lemma 1.3.5 we know that the only T-invariant measurable functions are constant functions.

Suppose f is constant. Then we know that $\int_X f \, d\mu = \mu_1(X) = 1$. But if $f = 1$ this implies that $\mu_1 = \mu$ which is impossible. So $\mu \notin M_{\text{erg}}(X, T)$. This completes the proof. □

In the remains of this section we will assume that X is a compact metric space, $\mathcal{B}$ is the Borel σ-algebra on X and T is a continuous transformation $T : X \to X$.

Before we go on with our proof, we have to define the structure of $M(X, \mathcal{B})$. We do this by defining a notion of convergence on $M(X, \mathcal{B})$.

The proof of the following lemma can be found in Appendix A.1.

Lemma 1.3.7. *For any $B \in \mathcal{B}$, and any $\mu \in M(X, \mathcal{B})$ we have*

$$\mu(B) = \sup_{C \subseteq B \mid C \text{ closed}} \mu(C) = \inf_{B \subseteq U \mid U \text{ open}}$$

Theorem 1.3.8. *Let $(X, \mathcal{B})$ be a measurable space. Let $\mu, \nu \in M(X, \mathcal{B})$. If*

$$\int_X f \, d\mu = \int_X f \, d\nu \text{ for all } f \in C(X)$$

Then $\mu = \nu$

Proof. By corollary 1.3.7, it is enough to show $\mu(C) = \nu(C)$ for all closed subsets C of X.

Let $\epsilon > 0$, by regularity of ν there exists an open set U_ϵ such that $C \subseteq U_\epsilon$ and $\nu(U_\epsilon \backslash C) < \epsilon$. We define the function $f \in C(X)$ as

$$f(x) = \frac{d(x, X \backslash U_\epsilon)}{d(x, X \backslash U_\epsilon) + d(x, C)}$$

We see that $\mathbb{1}_C \leq f \leq \mathbb{1}_{U_\epsilon}$. So also

$$\mu(C) \leq \int_X f \, d\mu = \int_X f \, d\nu \leq \nu(U_\epsilon) \leq \nu(C) + \epsilon$$

One can follow the same procedure to show that $\nu(C) \leq \mu(C) + \epsilon$. So $\mu(C) = \nu(C)$ for all closed subsets C of X. Hence, we can conclude that $\mu(B) = \nu(B)$ for all $B \in \mathcal{B}$ and therefore $\mu = \nu$. □

Now we can define what we mean with convergence in $M(X, \mathcal{B})$.

Definition 1.3.9 (Convergence in $M(X, \mathcal{B})$)**.** Let $(X, \mathcal{B})$ be a measurable space. A sequence $(\mu_n)_n$ converges to $\mu \in M(X, \mathcal{B})$ if and only if

$$\lim_{n \to \infty} \int_X f(x) \, d\mu_n(x) = \int_X f \, d\mu$$

for all $f \in C(X)$.

In the next part of this chapter we will proof that $M_{\text{inv}}(X, \mathcal{B}, \mu, T)$ fulfills all the requirements of the theorems proven above.

We will state the following theorem without proof. A proof can be found in *The Riesz representation theorem*[13].

Theorem 1.3.10 (Riesz-Markov-Kakutani Representation Theorem)**.** *Let X be a compact Hausdorff space. For any positive linear functional ψ on $C(X)$ there exists a unique Borel measure μ on X such that $\psi(f) = \int_X f(x) \, d\mu(x)$ for all $f \in C(X)$.*

Theorem 1.3.11. *The space $M(X, \mathcal{B})$ of all measures on $(X, \mathcal{B})$ is compact.*

Proof. Since X is compact, we know that $C(X)$ is seperable. Therefore there exists a countable and dense subset $\{f_n \mid n \in \mathbb{N}\} \subset C(X)$. Take (μ_n) a sequence in $M(X, \mathcal{B})$. By theorem 1.3.8, we know a sequence $(\mu_n) \subset M(X, \mathcal{B})$ converges to $\mu \in M(X, \mathcal{B})$ if and only if $\lim_{n\to\infty} \int_X f(x)\, d\mu_n(x) = \int_X f(x)\, d\mu(x)$ for all $f \in C(X)$. We will use this to proof that (μ_n) has a convergent subsequence.

Look at the sequence $(\int_X f_1\, d\mu_n)$. Since f is continuous on a compact space, we know that f has a maximal element on X. Since $(X, \mathcal{B}, \mu)$ is a probability space, we know $\int_X 1\, d\mu = 1$. So this sequence is bounded by $\|f_1\|_\infty$. We know by Bolzano Weierstrass that there exists a convergent subsequence of this sequence. Indicate this subsequence by $(\int_X f_1\, d\mu_{n_1})$. Now we look at f_2 and recognize that $(\int_X f_2\, d\mu_{n_1})$ is also bounded and therefore there exists a subsequence $(\int_X f_2\, d\mu_{n_2})$ which converges. If we continue this procedure we can create for each $i \in \mathbb{N}$ a subsequence (μ_{n_i}) of (μ_n) such that for each $j \leq i$ the sequence $(\int_X f_j\, d\mu_{n_i})$ converges. So if we take $i = n$, we know that the sequence $(\int_X f_j\, d\mu_{n_n})$ converges for every $j \in \mathbb{N}$.

Since $\{f_n \mid n \in \mathbb{N}\}$ is dense in $C(X)$ we know that $(\int_X f\, d\mu_{n_n})$ converges for all $f \in C(X)$.
Now define the functional

$$J : C(X) \to \mathbb{C}$$
$$J(f) = \lim_{n\to\infty} (\int_X f\, d\mu_{n_n})_n$$

We see that J is linear and continuous since the integral and the limit are continuous. Also $J(1) = 1$. By the Riesz-Markov-Kakutani Representation Theorem (1.3.10) we know there exists a $\mu \in M(X, \mathcal{B})$ such that $J(f) = \lim_{n\to\infty}(\int_X f\, d\mu_{n_n}) = \int_X f\, d\mu$ for all $f \in C(X)$. Therefore, the subsequence μ_{n_n} converges to $\mu \in M(X, \mathcal{B})$ and $M(X, \mathcal{B})$ is compact. □

Theorem 1.3.12. *The space $M_{inv}(X, \mathcal{B}, T)$ of T-invariant measures is compact.*

Proof. We will prove that $M_{\text{inv}}(X, \mathcal{B}, T)$ is closed in $M(X, \mathcal{B})$. Take $(\mu_n)_n \subset M_{\text{inv}}(X, \mathcal{B}, T)$ such that $\mu_n \to \mu \in M(X, \mathcal{B})$ for $n \to \infty$. Then for $f \in C(X)$ we have $\int_X f\, d\mu_n = \int_X f \circ T\, d\mu_n$. Since the integral is a continuous operator we can take the limits on both sides. This gives $\int_X f\, d\mu = \int_X f \circ T\, d\mu$ for all $f \in C(X)$. So $\mu \in M_{\text{inv}}(X, \mathcal{B}, T)$. Therefore, $M_{\text{inv}}(X, \mathcal{B}, T)$ is closed in $M(X, \mathcal{B})$. Since $M_{\text{inv}}(X, \mathcal{B}, T)$ is a closed subset of the compact space $M(X, \mathcal{B})$, we know that $M_{\text{inv}}(X, \mathcal{B}, T)$ is also compact. □

Theorem 1.3.13. *The space $M_{inv}(X, \mathcal{B}, T)$ of T-invariant measures is non-empty.*

Proof. Let $y \in X$. We define

$$\delta_y(x) = \begin{cases} 1 \text{ if } x = y \\ 0 \text{ else} \end{cases}$$

be a Dirac measure. Next we define $\mu_n \in M(X, \mathcal{B})$ by $\mu_n(A) = \frac{1}{n}\sum_{j=0}^{n-1}\delta_y(T^j A)$. This means that for $A \in \mathcal{B}$ we have $\mu_n(A) = \frac{1}{n}\left(\delta_y(A) + \delta_y(T^{-1}A) + ... + \delta_y(T^{-(n-1)}A)\right)$. Since $M(X, \mathcal{B})$ is compact, we know there exists a subsequence (μ_{n_k}), which converges to $\mu \in M(X, \mathcal{B})$ for $k \to \infty$. Now let $f \in C(X)$. Then

$$\begin{aligned} |\int_X f \cdot T \, d\mu - \int_X f \, d\mu| &= \lim_{k\to\infty} |\int_X f \cdot T \, d\mu_{n_k} - \int_X f \, d\mu_{n_k}| \\ &= \lim_{k\to\infty} |\frac{1}{n_k}\int_X \sum_{j=0}^{n_k-1}\left(f \cdot T^{j+1} - f \cdot T^j\right) \, d\delta_y| \end{aligned}$$

We recognize that we have a telescoping series. We can rewrite this as:

$$\begin{aligned} &= \lim_{k\to\infty} |\frac{1}{n_k}\int_X (f \cdot T^{n_k} - f) \, d\delta_y| \\ &\leq \lim_{k\to\infty} \frac{2\|f\|_\infty}{n_k} = 0 \end{aligned}$$

Therefore, $\mu \in M_{\text{inv}}(X, \mathcal{B}, T)$. So $M_{\text{inv}}(X, \mathcal{B}, T) \neq \emptyset$. □

Taking all of the above in account, we can now state the following proposition.

Proposition 1.3.14. *If $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$ is continuous on a compact metric space then $M_{erg}(X, \mathcal{B}, T) \neq \emptyset$.*

Proof. By the Krein-Milman Lemma(1.3.3) we know that $M_{\text{inv}}(X, \mathcal{B}, T)$ has an extreme point. By Theorem 1.3.6 we know that this point has to be an ergodic measure. Therefore $M_{\text{erg}}(X, \mathcal{B}, T) \neq \emptyset$. □

1.4 Birkhoff's Ergodic Theorem

We will first prove a lemma we will use in the proof of Birkhoff's Ergodic Theorem.

Lemma 1.4.1. *Let $(X, \mathcal{A}, \mu)$ be a measure space. Let $f : (X, \mathcal{A}, \mu) \to [0, \infty]$ be a measurable function and $E \in \mathcal{A}$ such that $\mu(E) = 0$. Then also $\int_E f \, d\mu = 0$.*

Proof. We know that the integral of f is given by

$$\int_E f \, d\mu = \sup\{\int_E s \, d\mu \mid s \text{ is a measurable function with } s \leq f\}$$

We know that a simple function can be written as $s = \sum_{i=1}^n a_i \chi_{E_i}$ with $E_i \in \mathcal{A}$, We recognize that $E \cap E_i \subseteq E$, so $\mu(E \cap E_i) \leq \mu(E) = 0$. Therefore $\int_E s \, d\mu = \sum_{i=1}^n a_i \mu(E \cap E_i) = 0$. So for all simple functions with $s \leq f$ we have $\int_E s \, d\mu = 0$. Therefore $\int_E f \, d\mu = 0$. □

Theorem 1.4.2 (Birkhoff's ergodic theorem)**.** *Let $(X, \mathcal{A}, \mu, T)$ be a measure preserving dynamical system. Let $f : (X, \mathcal{A}, \mu) \to \mathbb{R}$ be an element of $L^1(X, \mathcal{A}, \mu)$. Then the limit*

$$f_*(x) = \lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) \tag{2}$$

exists μ-almost everywhere. Moreover:

(i) $f_* \in L^1(X, \mathcal{A}, \mu)$.

(ii) f_* *is T-invariant.*

(iii) If $\mu \in M_{erg}$ *then* $\int_X f_* \, d\mu = \int_X f \, d\mu$

Proof. Take $f \in L^1(\mu)$. Then we can write f as $f = f^+ - f^-$ with f^+ and f^- both positive functions. Therefore without loss of generalization we can assume $f \geq 0$. We define $\overline{f}(x) := \limsup_{n\to\infty} f_n(x)$ and $\underline{f}(x) := \liminf_{n\to\infty} f_n(x)$.

We know that $f_n(x)$ converges at the points $x \in X$ where $\overline{f} \leq f \leq \underline{f}$. Instead of proving this we will proof the integral inequalities

$$\int_X \overline{f}(x) \, d\mu \leq \int_X f \, d\mu \leq \int_X \underline{f}(x) \, d\mu$$

This suffices since this would imply that $\overline{f} - \underline{f} \geq 0$ μ a.e. and therefore $\overline{f} = \underline{f} = f^*$ μ-a.e. This implies that f^* exists and is integrable μ-a.e..

We will assume that the left inequality is true and prove only the right inequality. If one wants to prove the left inequality, one can follow the the proof of the right inequality except for some minor details. We follow this procedure because in our proof of the right inequality we need the left inequality. One is able to prove the left inequality without the assumption that the right inequality is true.

Fix $\epsilon > 0$ arbitrary. We will show that

$$\int_X f \, d\mu \leq \int_X \underline{f} \, d\mu + \epsilon$$

By definition of the lim inf function we know that for each $x \in X$ there exists an $n \geq 1$ such that $f_n(x) \leq \underline{f}(x) + \epsilon$. Now we define the following function:

$$\tau : X \to \mathbb{N}$$
$$\tau(x) = min\Big\{ n \geq 1 \mid f_n(x) \leq \underline{f}(x) + \epsilon \Big\}$$

We know by assumption that $\overline{f} \in L^1(\mu)$ and therefore $\underline{f} \leq \overline{f} < \infty$ $\mu - a.e.$ Therefore, we know $\tau \in L^1(X, \mathcal{A}, \mu)$ and $\tau(x) < \infty$ $\mu - a.e.$ So for $M, N \in \mathbb{N}$ we have $\mu\{x \in X \mid f(x) > M\} \to 0$ and $\mu\{x \in X \mid \tau(x) > N\} \to 0$ for $N \to \infty$. So we know we can take M, N big enough so that $\mu\{x \in X \mid f(x) > M\} < \epsilon$ and $\mu\{x \in X \mid \tau(x) > N\} < \epsilon$. We define the set

$$E := \{x \in X \mid f(x) > M \text{ or } \tau(x) > N\}$$

We want to prove that $\int_E \underline{f}\, d\mu < \epsilon$

We know that for M, N sufficiently large we have $\mu(E) < \epsilon$. By lemma 1.4.1 we know that we have $\int_E \underline{f}\, d\mu < \epsilon$.

Now we change our function f so that we have no more "strange" points. We define

$$f'(x) = \begin{cases} f(x) \text{ if } x \notin E \\ min\{f(x), \underline{f}(x), M\} \text{ if } x \in E \end{cases}$$

We see that f' is bounded on E by M and by $\underline{f}$. Also $f' \leq f$ for $x \in X$.

Now define the sequence (f'_n) such that $f'_n = \frac{1}{n}\sum_{k=0}^{n-1} f'(T^k x)$. Then $f'_n \leq f_n$ and therefore the function

$$\tau' : X \to \mathbb{N}$$
$$\tau'(x) = min\left\{n \geq 1 \mid f'_n(x) \leq \underline{f}(x) + \epsilon\right\}$$

is everywhere finite and bounded by N. Also for $x \in E$ we know $\tau'(x) = 1$ because $f'(x) \leq \underline{f}(x)$. Now we have no more "bad" points.

Take $x \in X$. Decompose the orbit of x into pieces with initial points $x_0, x_1, ..., x_q$ so for $0 \leq j \leq q-1$ the orbit piece starting at x_j has length $\tau'(x_j)$. This means the average of f' along the orbit piece with first point x_j is smaller than $\underline{f}(x_j) + \epsilon \leq \underline{f}(x) + \epsilon$. This is possible for all orbit pieces except for the last one, which may not be long enough. We call the length of the last orbit piece $l \in \mathbb{N}$.

Now we can estimate the average of f' along the entire orbit of x. We calculate:

$$\begin{aligned}
\sum_{k=0}^{n-1} f(T^k x) &= \sum_{j=0}^{q-1} \sum_{j=0}^{\tau'(x_j)-1} f'(T^k x_j) + \sum_{k=0}^{l-1} f'(T^k x_q) \\
&\leq \sum_{j=0}^{q-1} \tau'(x_j) \left(\frac{1}{\tau'(x_j)} \sum_{k=0}^{\tau'(x_j)-1} f'(T^k x_j) \right) + lM \\
&\leq \sum_{j=0}^{q-1} \tau'(x_j) \left(\underline{f}(x) + \epsilon \right) + M \\
&\leq n(\underline{f}(x) + \epsilon) + NM
\end{aligned}$$

This gives us the final inequality:

$$f'_n(x) \leq \underline{f}(x) + \epsilon + \frac{NM}{n}$$

Integrating the left and right side and noting that $\mu(X) = 1$ gives:

$$\begin{aligned}
\int_X \frac{1}{n} \sum_{k=0}^{n-1} f'(T^k x)\, d\mu &\leq \int_X \underline{f}\, d\mu + \epsilon + \frac{NM}{n} \\
\frac{1}{n} \sum_{k=0}^{n-1} \int_X f'(T^k x)\, d\mu &\leq \int_X \underline{f}\, d\mu + \epsilon + \frac{NM}{n}
\end{aligned}$$

Because T is measure preserving this equals:

$$\begin{aligned}
\frac{1}{n} \sum_{k=0}^{n-1} \int_X f'(x)\, d\mu &\leq \int_X \underline{f}\, d\mu + \epsilon + \frac{NM}{n} \\
\int_X f'(x)\, d\mu &\leq \int_X \underline{f}\, d\mu + \epsilon + \frac{NM}{n}
\end{aligned}$$

Taking the limit $n \to \infty$ gives:

$$\int_X f'\, d\mu \leq \int_X \underline{f}\, d\mu + \epsilon$$

We know that $\int_X f'\, d\mu$ and $\int_X f\, d\mu$ only differ by ϵ (by definition of E). This gives us$\int_X f\, d\mu - \epsilon \leq \int_X f'\, d\mu \leq \int_X \underline{f}\, d\mu + \epsilon$. This gives us:

$$\int_X f\, d\mu \leq \int_X \underline{f}\, d\mu + 2\epsilon$$

Since this is true for all $\epsilon > 0$ we know that $\int_X f\, d\mu \leq \int_X \underline{f}\, d\mu$. This completes the first half of the proof. For the other inequality we want to prove, one can use the exact same method we used for the inequality just proven.

We prove that $\overline{f}$ is T-invariant. Take $x \in X$. We calculate:

$$\begin{aligned}\overline{f}(Tx) &= \limsup_{n\to\infty} \frac{f_n(Tx)}{n} \\ &= \limsup_{n\to\infty} \left(\frac{f_{n+1}(x)}{n+1} \cdot \frac{n+1}{n} - \frac{f(x)}{n} \right) \\ &= \limsup_{n\to\infty} \frac{f_{n+1}(x)}{n+1} = \overline{f}(x)\end{aligned}$$

Since $\overline{f}$ is T-invariant we know that f_* is T-invariant.

Assume $\mu \in M_{\text{erg}}$. Since f_* is T-invariant we know by lemma 1.3.5 that it must be constant μ-a.e. The integral over X equals $\int_X f \, d\mu$ so this must be the constant. □

The Birkhoff ergodic theorem is basically a generalisation of the Law of large numbers, which states the mean of independent identical distributed stochasts converges. We will show this now.

Example 1.4.3. Let $X = \mathbb{R} \times \mathbb{R} \times ...$ and $\mathcal{A} = B(\mathbb{R}) \times B(\mathbb{R}) \times ....$ Let μ be a measure on the measurable space $(\mathbb{R}, \mathcal{A})$. Take $\lambda = \mu \times \mu \times ...$ as the product measure. We let the transformation T be the left shift i.e. $T(x_0, x_1, ...) = (x_1, x_2,)$. Now we define the function

$$\begin{aligned} f : X &\to \mathbb{R} \\ f(x_0, x_1, ...) &= x_0 \end{aligned}$$

Now let $(X, \mathcal{A}, \lambda, T)$ be the corresponding dynamical system. We see that T is a measure preserving transformation since we use the product measure λ. This means we can use the Birkhoff ergodic theorem. This gives:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} x_j$$

We know by the Birkhoff Ergodic Theorem that this sequence converges. Since every x_i is identical and independently distributed, this is exactly the law of large numbers. Therefore, the Birkhoff ergodic theorem generalizes the law of large numbers. △

2 Entropy

This chapter is inspired by the following notes: *Introduction to Ergodic Theory*[4], *Ergodic Theory and Entropy*[1], *An Entropy Primer*[7] and *Entropy and Information Theory* [5].

Let $(X, \mathcal{A}, \mu)$ be a measure space. The main idea of entropy is well described by the following question:

Suppose we pick $x \in X$ at random. How much on average do we expect to know about the location of x learning that $x \in A \in \mathcal{A}$?

We will try to answer this question using the information function. Later on we will link this to entropy and prove some results. In the end of this chapter, we will see that we can use entropy to distinguish dynamical systems from one another.

2.1 Information function

Denote by $I(A)$ the amount of information gained learning that $x \in A \in \mathcal{A}$. Then the following requirements for $I : \mathcal{A} \to \mathbb{R}$ seem reasonable. Let $A, B \in \mathcal{A}$ then:

(i) $\mu(A) = \mu(B) \Rightarrow I(A) = I(B)$ (Information only depends on the size of the subset)

(ii) $\mu(A) = 1 \Rightarrow I(A) = 0$ (Knowing that $x \in A$ gives no extra information since A is a sure event)

(iii) If A and B are independent then $I(A \cap B) = I(A) + I(B)$

This motivates the following definition

Definition 2.1.1 (Information Function)**.** Let $(X, \mathcal{A}, \mu)$ be a probability space. Let $A \in \mathcal{A}$ be such that $\mu(A) > 0$. The function I defined by

$$\begin{aligned} I &: \mathcal{A} \to \mathbb{R} \\ I(A) &\mapsto -\log(\mu(A)) \end{aligned}$$

is called the information function.

This information function assigns to every element $A \in \mathcal{A}$ a non-negative real number.

Example 2.1.2. We will first look at the entropy of a transmission source which transmits symbols independently. We model this system as follows. Let $(X, \mathcal{A}, \mu, T)$ be a dynamical system with $X = \{a_1, a_2, ..., a_n\}^{\mathbb{N}}$ and $\mathcal{A}$ the σ-algebra generated by the cylinder sets i.e. $\mathcal{A} = \sigma(\Delta_n(a_{i_1}, ..., a_{i_n})) = \sigma(\{x \in X \mid x_{i_1} =$

$a_{i_1}, ..., x_{i_n} = a_{i_n}\}), \mu(\{x \in X \mid x_i = a_i\}) = p_i \ \forall i \in \{1, ..., n\}$ and T the left shift operator defined by

$$T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$$
$$T(a_1, a_2, ...) = (a_2, a_3, ...)$$

We define the entropy of this dynamical system by $H(p_1, ..., p_n) = -\sum_{i=1}^{n} p_i \log p_i$.

We note that the entropy defined above is basically the mean value of the information function over all possible symbols transmitted. To test whether this is a good definition, we look at the case where $p_i = 1$ for some $j \in \{1, ..., n\}$. In this case we have no uncertainty since we know only a_j can be transmitted. We check if our definition reflects this property. We calculate:

$$H(0, ..., 1, ...) = -\sum_{i=1}^{n} p_i \log p_1$$

With the convention that $0 \log 0 = 0$ we get:

$$H(0, ..., 1, ..) = 0$$

This indicates that our definition is correct.

Now we look at the case where our probabilities are uniformly distributed. If this is the case, our entropy should be maximal since we have no extra information about which symbols get transmitted more often. We will show that $H(p_1, ..., p_n)$ is maximal if we have an uniform distribution.

We define the function

$$\phi : X \to \mathbb{R}$$
$$x \mapsto -x \log x$$

We see that $\phi''(x) = -\frac{1}{x} < 0$ so ϕ is concave downward. Therefore by Jenssen's inequality we can calculate:

$$\begin{aligned}
\frac{1}{n} H(p_1, ..., p_n) &= \frac{1}{n} \sum_{i=1}^{n} \phi(p_i) \\
&\leq \phi\left(\frac{1}{n} \sum_{i=1}^{n} p_i\right) \\
&= \phi(\frac{1}{n}) \\
&= \frac{1}{n} \log n
\end{aligned}$$

This gives us that $H(\mu_1,,\mu_n) \leq \log n$.
Now we calculate:

$$\begin{aligned} H(\frac{1}{n},...,\frac{1}{n}) &= -\sum_{i=1}^{n} \frac{1}{n} \log \frac{1}{n} \\ &= \log n \end{aligned}$$

Therefore we know that the entropy is maximized for the uniform distribution.
△

In the example above we answered the main question of this chapter: *What is the average information gain per symbol?* Instead of an independent source we want to answer this question for a general dynamical system. To do this, we first need some knowledge about partitions.

Denote by $\Omega_X = \Omega_X(\mathcal{A})$ the space of all finite partitions on the measurable space $(X, \mathcal{A})$ consisting of measurable elements.

Definition 2.1.3 (Refinement of Partition)**.** Let $\alpha, \beta \in \Omega_X$. We say that β refines α if every $B \in \beta$ is included in some $A \in \alpha$. We denote $\alpha \leq \beta$

Definition 2.1.4 (Join of Partition)**.** Let $\alpha, \beta \in \Omega_X$. Then the join of α and β, denoted by $\alpha \vee \beta$, is the partition $X := \{A \cap B \mid A \in \alpha, B \in \beta\}$

Definition 2.1.5 (Traceback of Partition)**.** Let $\alpha \in \Omega_X$. Then the traceback of α, denoted by $T^{-1}\alpha$, is the partition $X := \{T^{-1}A \mid A \in \alpha\}$.

The join of two partitions will be a finer partition of the space X in the sense that for $\alpha, \beta \in \Omega_X$ we have $\alpha \vee \beta \geq \alpha$.

We will show in the following example that $\alpha \vee \beta$ and $T^{-1}\alpha$ are still partitions if $\alpha, \beta \in \Omega_X$.

Example 2.1.6. Let $\alpha, \beta \in \Omega_X$. We see that $\cup_{A\in\alpha} \cup_{B\in\beta} A \cap B = \cup_{A\in\alpha} \left(\cup_{B\in\beta} A \cap B\right)$. Since β is a partition, we know that this equals $\cup_{A\in\alpha} A = X$. So the union of all elements in $\alpha \vee \beta$ equals X. Now take $C_0, C_1 \in \alpha \vee \beta$. Then we can write $C_0 = A_p \cap B_r, C_1 = A_k \cap B_m$ for some $k, p \in \{1, ..., r\}$ and $m, r \in \{1, ..., l\}$. We calculate

$$\begin{aligned} C_0 \cap C_1 &= (A_p \cap B_r) \cap (A_k \cap B_m) \\ &= A_p \cap (B_r \cap A_k) \cap B_m \end{aligned}$$

Since the elements of α are disjoint, as well as the elements of β, we see that C_0 and C_1 are disjoint. So $\alpha \vee \beta$ is a partition.

Now look at $T^{-1}\alpha$. We know $\cup_i A_i = X$ and $T^{-1}X = X$, hence $\cup_i T^{-1}A_i = T^{-1} \cup_i A_i = T^{-1}X = X$

Take $B_0, B_1 \in T^{-1}\alpha$. Then $B_0 = T^{-1}A_0$ and $B_1 = T^{-1}A_1$ for some $A_0, A_1 \in \alpha$. So we have $B_0 \cap B_1 = T^{-1}A_0 \cap T^{-1}A_1 = \{x \in X \mid Tx \in A_0\} \cap \{x \in X \mid Tx \in A_1\}$. Since $A_0 \cap A_1 = \emptyset$ the set above is also empty because there exists no $x \in X$ such that $x \in A_0 \cap A_1$.

So $T^{-1}\alpha$ is a partition △

2.2 Entropy

In this section we will mathematically define what entropy is and prove some results regarding the properties of entropy. In the last part of this section we will prove that we can distinguish dynamical systems by looking at the entropy.

Definition 2.2.1 (Entropy w.r.t. a partition)**.** Let $(X, \mathcal{A}, \mu)$ be a probability space. Let $\alpha \in \Omega_X$ where α is the partition $X = \cup_{j=1}^{r} A_j$. We define the entropy with respect to α by

$$H(\alpha) = -\sum_{j=1}^{r} \mu(A_j) \log \mu(A_j)$$

What this function intuitively does, is assigning a number to the amount of "uncertainty" in our dynamical system.

Definition 2.2.2 (Conditional Entropy)**.** Let $\alpha, \beta \in \Omega_X$. We define the conditional entropy of α given β by

$$H(\alpha \mid \beta) = -\sum_{A \in \alpha} \sum_{B \in \beta} \log \left(\frac{\mu(A \cap B)}{\mu(B)} \right) \mu(A \cap B)$$

Intuitively, the conditional entropy can be seen as the average uncertainty one has about the location of x knowing that x lies in $B \in \beta$.

Example 2.2.3. Let $\alpha = \{A_1, ..., A_r\}$ and $\beta = \{B_1, ..., B_s\}$ be elements of Ω_X. Let μ_{B_k} be the conditional measure $\mu_{B_k} = \mu(\cdot \mid B_k)$. To illustrate what conditional entropy is, we make the following calculation:

$$\begin{aligned} H_\mu(\alpha \mid \beta) &= -\sum_{j=1}^{r}\sum_{k=1}^{s}\mu(A_j\cap B_k)\log\mu(A_j\cap B_k) + \sum_{k=1}^{s}\mu(B_k)\log\mu(B_k) \\ &= -\sum_{j=1}^{r}\sum_{k=1}^{s}\mu(A_j\cap B_k)\log\mu(A_j\cap B_k) + \sum_{j=1}^{r}\sum_{k=1}^{s}\mu(A_j\cap B_k)\log\mu(B_k) \\ &= -\sum_{k=1}^{s}\sum_{j=1}^{r}\big(\mu(A_j\cap B_k)\big(\log\mu(A_j\cap B_k) - \log\mu(B_k)\big)\big) \\ &= -\sum_{k=1}^{s}\mu(B_k)\sum_{j=1}^{r}\frac{\mu(A_j\cap B_k)}{\mu(B_k)}\log\frac{\mu(A_j\cap B_k)}{\mu(B_k)} \\ &= \sum_{k=1}^{s}\mu(B_k)H_{\mu_{B_k}}(\alpha) \end{aligned}$$

We see conditional entropy can be seen as the mean value of $H(\alpha)$ knowing we have to condition on β. △

Proposition 2.2.4. *Let $(X, \mathcal{A}, \mu, T)$ be a measure preserving dynamical system. Let $\alpha, \beta, \gamma \in \Omega_X$. Then:*

(i) $H(T^{-1}\alpha) = H(\alpha)$

(ii) $H(\alpha \vee \beta) = H(\alpha) + H(\beta \mid \alpha)$

(iii) $H(\beta \mid \alpha) \leq H(\beta)$

(iv) $H(\alpha \vee \beta) \leq H(\alpha) + H(\beta)$

(v) If $\alpha \leq \beta$ then $H(\alpha) \leq H(\beta)$

(vi) $H(\alpha \vee \beta \mid \gamma) = H(\alpha \mid \gamma) + H(\beta \mid \alpha \vee \gamma)$

(vii) If $\beta \leq \alpha$ then $H(\gamma \mid \alpha) \leq H(\gamma \mid \beta)$

(viii) If $\beta \leq \alpha$ then $H(\beta \mid \alpha) = 0$

To proof of these identities can be found in Appendix A.2 at the end of this thesis.

Let $(X, \mathcal{A}, \mu, T)$ be a dynamical system. We define the orbit of $x \in X$ as the sequence $(x, Tx, T^2x, ...)$. What we do with this definition, is looking at some $x \in X$ and register where x is mapped to by T.

We will first state the definition of entropy of a measure preserving transformation with respect to a partition, and later we will explain why this definition makes sense.

Definition 2.2.5 (Entropy w.r.t. a measure)**.** Let $(X, \mathcal{A}, \mu, T)$ be a dynamical system. We define the entropy of T with respect to μ and $\alpha \in \Omega_X$ as

$$h_\mu(T; \alpha) := \lim_{n\to\infty} \frac{1}{n} H_\mu(\bigvee_{i=0}^{n-1} T^{-i}\alpha)$$

First, we look at the elements in $\bigvee_{i=0}^{n-1} T^{-i}\alpha$. If we follow definition 2.1.4 we note that elements of this partition are of the form $A_{i_0} \cap T^{-1}A_{i_1} \cap ... \cap T^{-(n-1)}A_{i_{n-1}}$. Elements in these sets are $x \in X$ such that $x \in A_{i_0}, Tx \in A_{i_1}, ..., T^{n-1}x \in A_{i_{n-1}}$. What this definition does, is measuring the average information gain we get from knowing an element in the orbit of $x \in X$. This is equivalent with what we did with the transmission source at the beginning of this chapter.

Lemma 2.2.6. *Let (a_n) be a subadditive sequence i.e. $a_{n+p} \le a_n + a_p$ for all $n, p \in \mathbb{N}$. Then the limit $\lim_{n\to\infty} \frac{a_n}{n}$ exists.*

Proof. First we fix $k \in \mathbb{N}$ arbitrary. Then subadditivity implies $a_{nk} \le na_k$ for all $n \ge 1$. So also $\frac{a_{nk}}{k} \le \frac{na_k}{k} \le \frac{a_k}{k}$ for all $n \ge 1$. Take $m \in \mathbb{N}$. Then $m = nk + l$ for some $l \in \{0, 1, ..., k-1\}$. We show:

$$\frac{a_m}{m} \le \frac{a_{nk}}{nk+l} + \frac{a_l}{nk+l} \le \frac{a_k}{k} + \frac{a_l}{nk+l} \le \frac{a_k}{k} + \frac{a_l}{nk}$$

If we take $m \to \infty$ then also $n \to \infty$. Therefore

$$\limsup_{m\to\infty} \frac{a_m}{m} \le \frac{a_k}{k} \text{ for all } k \ge 1$$

So also

$$\limsup_{m\to\infty} \frac{a_m}{m} \le \liminf_{k\to\infty} \frac{a_k}{k}$$

Therefore, the limit $\lim\limits_{n\to\infty} \frac{a_n}{n}$ exists. □

Lemma 2.2.7. *Let $(X, \mathcal{A}, \mu, T)$ be a measure preserving system and $\alpha \in \Omega_X$. Then $\lim\limits_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}\alpha)$ exists.*

Proof. Let $a_n = H(\bigvee_{i=0}^{n-1} T^{-i}\alpha)$. Then by Lemma A.2.1 we can calculate:

$$\begin{aligned} a_{n+p} &= H(\bigvee_{i=0}^{n+p-1} T^{-i}\alpha) \\ &\le H(\bigvee_{i=0}^{n-1} T^{-i}\alpha) + H(\bigvee_{i=n}^{n+p-1} T^{-i}\alpha) \\ &= a_n + H(\bigvee_{i=0}^{p-1} T^{-i}\alpha) \\ &= a_n + a_p \end{aligned}$$

So the sequence (a_n) is subadditive. By Lemma 2.2.6 we know that the limit $\lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}\alpha)$ exists. □

Definition 2.2.8 (Measure theoretical entropy)**.** Let $(X, \mathcal{A}, \mu, T)$ be a measure preserving dynamical system. We define the measure theoretical entropy of T with respect to μ as

$$h_\mu(T) := \sup_{\alpha \in \Omega_X} h_\mu(T;\alpha)$$

Definition 2.2.9 (Measure-Theoretical Isomorphism)**.** Let $(X, \mathcal{A}, \mu, T)$ and $(Y, \mathcal{B}, \nu, S)$ be dynamical systems. Let $\psi : (X, \mathcal{A}, \mu, T) \to (Y, \mathcal{B}, \nu, S)$ be a map such that ψ is measurable, measure preserving, invertible and for almost all $x \in X$ $\psi(Tx) = S(\psi(x))$; that is, such that the diagram

$$\begin{array}{ccc} X & \xrightarrow{T} & X \\ \psi\downarrow & & \downarrow\psi \\ Y & \xrightarrow{S} & Y \end{array}$$

commutes almost everywhere. Then we call ψ an isomorphism.

Theorem 2.2.10. *Entropy is invariant under measure theoretical isomorphism.*

Proof. Let $(X, \mathcal{A}, \mu, T)$ and $(Y, \mathcal{B}, \nu, S)$ be two measure preserving dynamical systems which are isomorphic with isomorphism $\psi : (X, \mathcal{A}, \mu, T) \to (Y, \mathcal{B}, \nu, S)$. Let $\beta \in \Omega_Y$ and take $\alpha = \psi^{-1}\beta$. We will proof that the entropy of T with respect to α is always greater then or equal to the entropy of S with respect to β.

Take $B_{i_0}, ..., B_{i_{n-1}} \in \beta$. Then:

$$\begin{aligned} \nu\left(B_{i_0} \cap S^{-1}B_{i_1} \cap ... \cap S^{-(n-1)}B_{i_{n-1}}\right) &= \mu\left(\psi^{-1}B_{i_0} \cap \psi^{-1}S^{-1}B_{i_1} \cap ... \cap \psi^{-1}S^{-(n-1)}B_{i_{n-1}}\right) \\ &= \mu\left(\psi^{-1}B_{i_0} \cap T^{-1}\psi^{-1}B_{i_1} \cap ... \cap T^{-(n-1)}\psi^{-1}B_{i_{n-1}}\right) \\ &= \mu\left(A_{i_0} \cap T^{-1}A_{i_1} \cap ... \cap T^{-(n-1)}A_{i_{n-1}}\right) \end{aligned}$$

We define

$$\begin{aligned} A(n) &= A_{i_0} \cap T^{-1}A_{i_1} \cap ... \cap T^{-(n-1)}A_{i_{n-1}} \\ B(n) &= B_{i_0} \cap S^{-1}B_{i_1} \cap ... \cap S^{-(n-1)}B_{i_{n-1}} \end{aligned}$$

With this work done, we can now make our calculation.

$$\begin{aligned}
h_\nu(S) &= \sup_{\beta\in\Omega_Y} h_\nu(\beta, S) \\
&= \sup_{\beta\in\Omega_Y} \lim_{n\to\infty} \frac{1}{n} H_\nu\left(\bigvee_{i=0}^{n-1} S^{-i}\beta\right) \\
&= \sup_{\beta\in\Omega_Y} \lim_{n\to\infty} \left(-\frac{1}{n} \sum_{B(n)\in\vee_{i=0}^{n-1} S^{-i}\beta} \nu(B(n)) \log \nu(B(n))\right) \\
&= \sup_{\psi^{-1}\beta} \lim_{n\to\infty} \left(-\frac{1}{n} \sum_{A(n)\in\vee_{i=0}^{n-1} T^{-i}\psi^{-1}\beta} \mu(A(n)) \log \mu(A(n))\right) \\
&= \sup_{\psi^{-1}\beta} h_\mu\left(\psi^{-1}\beta, T\right) \\
&\leq \sup_{\alpha\in\Omega_X} h_\mu(\alpha, T) = h_\mu(T)
\end{aligned}$$

The proof of $h_\mu(T) \leq h_\nu(S)$ can be similarly done. Therefore, we have $h_\mu(T) = h_\nu(S)$ and we can conclude that entropy is invariant under measure theoretical isomorphism. □

Example 2.2.11. Let $X = \{0,1\}^{\mathbb{N}}$ endowed with the σ-algebra $\mathcal{A}$ generated by the cylinder sets in X i.e. $\mathcal{A} = \sigma(\{(x_1, ..., x_n) \mid x_i \in \{0,1\}\ \forall 1 \leq i \leq n\})$. Let T be the left shift and μ defined by $\mu(x_1, ..., x_n) = 2^{-n}$. Let $(X, \mathcal{A}, \mu, T)$ denote the corresponding dynamical system.

Let $[0,1)$ be endowed with the Borel σ-algebra on $[0,1)$. Let λ denote the Lebesgue measure and let $S : ([0,1), B(\mathbb{R}), \lambda) \to ([0,1), B(\mathbb{R}), \lambda)$ be the doubling map

$$\begin{aligned}
S : ([0,1), B(\mathbb{R}), \lambda) &\to ([0,1), B(\mathbb{R}), \lambda) \\
x &\mapsto 2x \ mod\ 1
\end{aligned}$$

Let $([0,1), B(\mathbb{R}), \lambda, S)$ denote the corresponding dynamical system.

We will prove that $(X, \mathcal{A}, \mu, T)$ and $([0,1), B([0,1)), \lambda, S)$ are isomorphic.

We define $X' = \left\{x \in X \mid \nexists n\ s.t.\ x_m = 1\ for\ m \geq n\right\}$ and $X'_n = \left\{x \in \{0,1\}^{\mathbb{N}} \mid x_m = 1\ \forall m \geq n\right\}$. We calculate

$$\mu(X \backslash X') = \mu\left(\bigcup_{n=1}^{\infty} X'_n\right)$$

Since this sequence is monotone increasing we have

$$= \lim_{n\to\infty} \mu\left(X'_n\right) = \lim_{n\to\infty} 2^{-n} = 0$$

So the set $\left\{x \in \{0,1\}^{\mathbb{N}} \mid \exists n\ s.t.\ x_m = 1\ \forall m \geq n\right\}$ is negligible.

We define the map

$$\psi : (X', \mathcal{A}, \mu, T) \to ([0,1), B([0,1)), \lambda, S)$$
$$(x_1, ...) \mapsto \sum_{i=1}^{n-1} 2^{-i} x_i$$

To show that ψ is measure preserving, we prove that it is measure preserving for the cylinder sets, which generate the σ-algebra on X. Let $C(a_1, ..., a_n) = \{x \in X \mid x_1 = a_1, ..., x_n = a_n\}$ be a cylinder set of length n. We know that $\mu(C(a_1, ..., a_n)) = 2^{-n}$. We calculate:

$$\psi(C(a_1, ..., a_n)) = [\sum_{i=1}^{n-1} 2^{-i} a_i, \sum_{i=1}^{n-1} 2^{-i} a_i + \sum_{i=n+1}^{\infty} 2^{-i}) = [\sum_{i=1}^{n-1} 2^{-i} a_i, \sum_{i=1}^{n-1} 2^{-i} a_i + 2^{-n})$$

So with the Lebesgue measure we have $\lambda(\psi(C(a_1, ..., a_n))) = 2^{-n}$. So ψ is measure preserving.

Since every real number in $[0,1)$ has an unique binary expansion, we know that ψ is measurable and invertible.

Let $(a_1, ..., a_n, ...) \in C(a_1, ..., a_n)$. We calculate

$$\begin{aligned}
\psi(T(a_1, ..., a_n, ...)) &= \psi((a_2, ..., a_n, ...)) \\
&= [\sum_{i=2}^{n} 2^{-(i-1)} a_i, \sum_{i=2}^{n} 2^{-(i-1)} a_i + \sum_{i=n}^{\infty} 2^{-i}] \\
&= [\sum_{i=2}^{n} 2^{-(i-1)} a_i, \sum_{i=2}^{n} 2^{-(i-1)} a_i + 2^{-n+1}] \\
S(\psi(a_1, ..., a_n, ...)) &= S([\sum_{i=1}^{n} 2^{-i} a_i, \sum_{i=1}^{n} 2^{-i} a_i + \sum_{i=n+1}^{\infty} 2^{-i}]) \\
&= [\sum_{i=1}^{n} 2^{-(i-1)} a_i\ mod 1, \left(\sum_{i=1}^{n} 2^{-(i-1)} a_i + 2^{-n+1}\right)\ mod 1) \\
&= [\sum_{i=2}^{n} 2^{-(i-1)} a_i, \sum_{i=2}^{n} 2^{-(i-1)} a_i + 2^{-n+1})
\end{aligned}$$

So $\psi T = S\psi\ \mu$ a.e.

Since all the requirements of an isomorphism are fulfilled, we know that $(X, \mathcal{A}, \mu, T)$ and $([0,1), B([0,1)), \lambda, S)$ are isomorphic. △

We will now state a theorem we will use to calculate the entropy of a Bernoulli(1/2, 1/2) shift. A proof of this theorem can be found in *Notes for the Course: Ergodic Theory and Entropy*[1], page 98-99.

Definition 2.2.12 (Generator)**.** We call a partition $\alpha \in \Omega_X(\mathcal{A})$ a generator with respect to T if $\sigma\left(\bigvee_{i=0}^{\infty} T^{-i}\alpha\right) = \mathcal{A}$

Theorem 2.2.13 (Kolmogorov and Sinai)**.** *Let $(X, \mathcal{A}, \mu, T)$ be a measure preserving dynamical system. If $\alpha \in \Omega_X(\mathcal{A})$ is a generator with respect to T and $H(\alpha) < \infty$, then $h_\mu(T) = h_\mu(T, \alpha)$.*

Example 2.2.14 (Entropy of the Bernoulli(1/2, 1/2) shift)**.** Let $(X, \mathcal{A}, \mu, T)$ be the Bernoulli(1/2, 1/2) shift as in Example 2.2.11. We want to calculate the entropy of this dynamical system.

We define $\alpha \in \Omega_X$ as $\alpha = \{A_0, A_1\}$ with $A_i = \{x \in X \mid x_0 = i\}$. Then we recognize that

$$T^{-n}A_i = \{x \in X \mid x_n = i\}$$

Now we note that elements in $\bigvee_{i=0}^{n-1} T^{-i}\alpha$ are of the form

$$A_{i_0} \cap A_{i_1} \cap ... \cap T^{-(n)}A_{i_n} = \{x \in X \mid x_0 = i_0, ..., x_n = i_n\}$$

However, these are exactly the cylinder sets of length n. By definition of $\mathcal{A}$ we know these sets generate $\mathcal{A}$. So α is a generator with respect to T. We see that

$$H(\alpha) = H(\{A_0, A_1\}) = -\left(\frac{1}{2}\log(\frac{1}{2}) + \frac{1}{2}\log(\frac{1}{2})\right) = \log(2) < \infty$$

So we can use Theorem 2.2.13 to calculate $h_\mu(T)$.

We know that $\{A_0, A_1\}, T^{-1}\{A_0, A_1\}, ..., T^{-(n-1)}\{A_0, A_1\}$ are independent since all entries of $x \in X$ follow the Bernoulli(1/2, 1/2) distribution. Therefore $\mu(T^{-i}\alpha) = \mu(\alpha) = (1/2, 1/2)$. So we calculate

$$\begin{aligned} H(\alpha \vee T^{-1}\alpha \vee ... \vee T^{-(n-1)}\alpha) &= H(\alpha) + H(T^{-1}\alpha) + ... + H(T^{-(n-1)}\alpha) \\ &= nH(\alpha) = \log(2) \end{aligned}$$

We can now calculate

$$\begin{aligned} h_\mu(T) &= \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} \alpha) \\ &= \lim_{n\to\infty} \frac{n}{n} \log(2) = \log(2) \end{aligned}$$

△

Example 2.2.15 (Entropy of the Double Angle Map)**.** Let $([0,1), B([0,1)), \lambda, S)$ be the Double Angle Map as in Example 2.2.11. Since we have proven in Example 2.2.11 that the Double Angle Map and the Bernoulli(1/2, 1/2) shift are isomorphic and we know that entropy is invariant under isomorphism (Theorem 2.2.10) we can directly conclude that $h(S) = \log(2)$. △

In this chapter we have studied entropy in the discrete sense. In the next chapter we will use entropy in the differential sense. Therefore, we will now state the definition of entropy we will use in the next chapter

Definition 2.2.16 (Differential Entropy)**.** Let X be a random variable with probability density function f. Denote the support of X by χ. We define the differential entropy H of X to be

$$H(X) = -\int_{\chi} f(x) \log f(x)\, dx = -\mathbb{E}\left(\log f(x)\right)$$

3 Entropy in financial mathematics

This chapter is strongly based on *The Entropy Theory of Stock Option Pricing*[6].

The goal of this chapter is to find a model for the valuation of European options on the stock market. We will do this by finding the probability density of a stock which results in the fair price of our European option. By fair, we mean that whichever price our stock will have at expiration time, we will always be able to pay off this option at expiration time by using the right buying/selling strategy of the underlying stock. We call the probability density with this property the risk neutral probability density. We will find this risk neutral probability density by finding the probability density that results in the highest entropy, subject to some boundary conditions. With this probability density, we will then derive the price of a call and a put option on this stock process. It turns out that the maximum entropy probability density is given by the gamma density. We will derive the parameters of this gamma density by looking at boundary conditions.

The idea behind this derivation is that we want to find an alternative model to the Black and Scholes model. Instead of the lognormal distribution one obtains in the derivation of the Black and Scholes model, we will now find a gamma distribution. The improvement here is that in the derivation of the gamma distribution we do not assume that the interest is constant over time. Therefore, we need less assumptions then the Black and Scholes model.

First, we will derive the probability density for a stock process. This turns out to be the easy case of the probability density we derive later when looking at an European option that expires at time T with strike price K. To derive the density in the latter case, we assume a stock follows a square-root process. We will solve the Kolmogorov equation of this process and find that in the limit this solution gives the gamma density.

3.1 Motivation for using entropy

We assume that we are working in an information efficient market. This means every player in our market has access to the same information. We also assume we are working in a market with many players, whom which every player is competitive, i.e. wants to make money.

The price of a stock in a competitive market is a reflection of the belief that players in our market have in this stock. This means if more than 50 percent of our players think the price will move up, this will indeed happen, until the balance is 50-50 again. Therefore, this equilibrium state reflects the state of maximum uncertainty about the stock movement. Notice that we are talking about a maximum ignorance market belief. The belief of individual investors can be diverse.

We are looking for the probability density that reflects this maximum uncertainty state. As we have seen in Chapter 2, a good mathematical construction for measuring uncertainty of a system is entropy.

In this chapter, we will derive the probability density that has maximal entropy and therefore is the best probability density we can use for evaluating our stock.

3.2 Risk neutral probability density for stock process

To find the probability density of a stock without cash flow, we want to solve the following system of equations.

Proposition 3.2.1. *Let $S(t)$ be the stock price at time t and let $\mathcal{P}$ be the possible values of $S(t)$. Let $P(t)$ be the price of a bond at time t that pays 1 at time T and let σ^2 be the standard deviation of $S(t)$.*

If we have full knowledge of the information set $I = \{\mathcal{P}, S(0), P(0), \sigma^2\}$, then the risk-neutral density $f(S(T))$ solves the following maximization problem

$$\max_{g \in M(\mathcal{P})} - \int_{\mathcal{P}} g \log g \, dS(T)$$

with $M(\mathcal{P})$ the space of all probability measures with support $\mathcal{P}$

Subject to boundary conditions

$$\int_{\mathcal{P}} g \, dS(T) = 1 \tag{3}$$

$$\mathbb{E}_g(S_T) = \int_{\mathcal{P}} g S(T) \, dS(T) = S(0)/P(0) \tag{4}$$

$$\mathbb{E}_g(S_T^2) = \int_{\mathcal{P}} g S(T)^2 \, dS(T) = \sigma^2 + (S(0)/P(0))^2 \tag{5}$$

$$\tag{6}$$

Boundary condition 3 ensures our density is a probability density. We need this because we are dealing with an economical system subject to probability.

Boundary condition 4 ensures the no-arbitrage condition and therefore also the risk-neutrality of our probability density. Assume $\mathbb{E}_g(S(T)) > S(0)/P(0)$ then we can borrow $S(0)$ and buy stock with value $S(0)$ at time $t = 0$. Together with interest we have to pay back $S(0)/P(0)$ at time $t = T$. Since the expectation value of our stock is bigger than $S(0)/P(0)$ we can always sell our stock, pay back the loan and still have money left. This means arbitrage.

Assume $\mathbb{E}_g(S(T)) < S(0)/P(0)$. Then we short sell stock with value $S(0)$ at time $t = 0$ and put this in the bank. Together with interest this will grow to $S(0)/P(0)$ at time $t = T$. Since the expectation value of our stock is smaller then $S(0)/P(0)$ we can always sell back our stock and still have money left. This means arbitrage. Therefore, Boundary condition 4 is the only possibility that ensures no-arbitrage.

Boundary condition 5 ensures that the variance of $S(T)$ is still σ^2.

Proposition 3.2.2. *The unique solution $f(S(T))$ that solves Proposition 3.2.1 is*

$$f(S(T)) = \frac{\exp(\lambda_1 S(T) + \lambda_2 S(T)^2)}{\int_{\mathcal{P}} \exp(\lambda_1 S(T) + \lambda_2 S(T)^2) dS(T)}$$

where λ_1 and λ_2 are chosen so that $f(S(T))$ satisfies the boundary conditions of Assumption 3.2.1.

Proof. To find the extremum of this functional we will use the Lagrange multiplier method.

The Lagrange multiplier method tells us that to solve Proposition 3.2.1, we can also find the extremum of

$$\begin{aligned}\mathcal{L}(f, S(T)) = \int_{\mathcal{P}} \left(-f \log f + \lambda_0 f + \lambda_1 f S(T) + \lambda_2 f S(T)^2 \right) dS(T) \\ - 1 - S(0)/P(0) - \sigma^2 - (S(0)/P(0))^2\end{aligned}$$

Finding the extremum of this is the same as finding the extremum of

$$F(f, S(T)) = -f \log f + \lambda_0 f + \lambda_1 f S(T) + \lambda_2 f S(T)^2$$

To find the extremum of this functional we can solve the Euler-Lagrange equation

$$0 = \frac{\partial F}{\partial f} + \frac{d}{dS(T)} \left(\frac{\partial F}{\partial f'} \right)$$

Since our functional is independent of f' this equation simplifies to

$$\begin{aligned} 0 &= -\log f - 1 + \lambda_0 + \lambda_1 S(T) + \lambda_2 S(T)^2 \\ f(S(T)) &= \exp\{\lambda_1 S(T) + \lambda_2 S(T)^2\} \end{aligned}$$

To satisfy constraint 1 of our system we have to normalize f. So

$$f(S(T)) = \frac{\exp\{\lambda_1 S(T) + \lambda_2 S(T)^2\}}{\int_{\mathcal{P}} \exp\{\lambda_1 S(T) + \lambda_2 S(T)^2\} dS(T)}$$

where we choose λ_1 and λ_2 such that our function fulfills constraints 4 and 5.

We know that the function $g \to g\log(g)$ is convex. The space of all probability densities on $\mathcal{P}$ is convex. Therefore the functional $\int g \log g \; dS(T)$ is convex. So the functional $-\int g \log g \; dS(T)$ is concave. Therefore $f(S(T))$ is the global maximum. For a rigorous proof of this, look at *Optimization by Vector Space Methods*[10], page 191. □

3.3 Risk neutral probability density for stock process under cash-flow assumption

Assume we have an option with expiration date T which pays no dividend at times $t < T$ but pays dividend d_j at times $j > T$. We consider this dividend stream $A_{t,T} = \{d_j \mid t \geq j > T\}$ from time T to time t.

Now consider the shifted stream $B_{t,K} = \{d_{j-(K-t)} \mid t \geq j > K\}$, which is the stream $A_{t,T}$ shifted $K - t$ to the left. Now take $B_{T,0} = \{d_{j-T} \mid T \geq j > 0\}$ which is exactly the stream $A_{t,T}$ shifted to start at $t = 0$. We will determine the price $S(T)$ of our stock by looking at the random price $x(T)$ of the cash-flow stock $B_{0,T}$, Intuitively this $x(t)$ will be low right after a dividend payment and high right before one. Let $x(0) = x_0$.

We see that $B_{0,T} = A_{t,T}$ and therefore $X(T) = S(T)$. Therefore, they also have identical probability distributions at time $t = T$. We assume that $x(t)$ follows a plausible stochastic process, and use this to determine the probability distribution of $x(T)$.

Assumption 3.3.1. *The price $x(t)$ of a constant cash-flow security changes on $[0, \infty)$ according to*

$$dx = (-bx + c)dt + \sqrt{2ax}\, dz$$

where dz is a standard Brownian motion with $\mathbb{E}(dz) = 0$ and $Var(dz) = dt$. The parameters $a, b, c > 0$ are constant. This is called the square-root process.

The coefficient $(-bx + c)$ gives us a mean reversion. If $x > \frac{c}{b}$ then this coefficient becomes smaller than zero. If $x < \frac{c}{b}$ then this coefficient becomes bigger then zero. Therefore, this process will always fluctuate around its mean $\frac{c}{b}$. We see that as $x(t)$ deviates more from the mean, the standard deviation term $\sqrt{2ax}\, dx$ becomes bigger and so $x(t)$ returns faster to the mean. This is a process often used in finance and is also known as the Cox-Ingersoll-Ross model.

Proof of the following lemma can be found in any good book on stochastic processes. For example in the book *Stochastic Processes*[11].

Lemma 3.3.2 (Kolmogorov Backward Equation)**.** *Let $W(t)$ be a Brownian motion. Assume a system $y(t)$ evolves according to*

$$dy(t) = \mu(y(t), t)dt + \sigma(y(t), t)dW(t)$$

Then the probability density $p(y, t)$ of $y(t)$ is given by the equation

$$\frac{\partial}{\partial t} p(y, t) = \mu(y, t) \frac{\partial}{\partial y} p(y, t) + \frac{1}{2}\sigma^2(y, t) \frac{\partial^2}{\partial y^2} p(y, t) \tag{7}$$

Let $g(x, t \mid x(0) = x_0, t = 0)$ be the probability density of the random variable $x(t)$ at time $t > 0$ with the boundary condition $x(0) = x_0$ at time $t = 0$. We will find an expression for $g(x, t \mid x_0, 0)$ by solving the Kolmogorov backward equation provided by lemma 3.3.2. This gives us the following equation:

$$g_t = (-bx + c)g_x + \frac{1}{2}(\sqrt{2ax})^2 g_{xx}$$
$$g_t = (-bc + c)g_x + axg_{xx}$$

This gives us an extra equation we add to Proposition 3.2.1. This leads to the following system.

Proposition 3.3.3. *If we have full knowledge of the information set $I = \{\mathcal{P}, S(0), P(0), \sigma^2\}$, then the risk-neutral density $f(S(T))$ solves the following maximalization problem*

$$\max_{g \in M(\mathcal{P})} - \int_{\mathcal{P}} g \log g \; dS(T) \tag{8}$$

with $M(\mathcal{P})$ the space of all probability measures with support $\mathcal{P}$

Subject to boundary conditions:

$$\int_{\mathcal{P}} g \; dS(T) = 1 \tag{9}$$

$$\mathbb{E}_g(S(T)) = \int_{\mathcal{P}} gS(T) \; dS(T) = S(0)/P(0) \tag{10}$$

$$\mathbb{E}_g(S(T)^2) = \int_{\mathcal{P}} gS(T)^2 \; dS(T) = \sigma^2 + (S(0)/P(0))^2 \tag{11}$$

$$axg_{xx} + (-bx + c)g_x = g_t \tag{12}$$

$$\tag{13}$$

Proposition 3.3.4. *The unique solution to the program in Proposition 3.3.3 is the gamma density*

$$f(S(T)) = \gamma(S(T) \mid u, v) = \frac{u^v}{\Gamma(v)} S(T)^{v-1} e^{-uS(T)}$$

where

$$u = \frac{(S(0)/P(0))}{\sigma^2}$$
$$v = \frac{(S(0)/P(0))^2}{\sigma^2}$$

Proof. We will state the solution of Equation 12. If one wants to see the derivation of this solution, one can look at the book *A Second Course in Stochastic Processes* [9] page 334.

The solution to equation 12 is given by

$$g(x,t \mid x_0, 0) = \left(\frac{b}{a}\right)^{c/a} x^{c/a-1} e^{-\frac{b}{a}x} \sum_{n=0}^{\infty} e^{-nat} \left(\frac{\Gamma(n+1)}{\Gamma(n+c/a)}\right) L_n^{(c/a-1)}\left(\frac{b}{a}x_0\right) L_n^{(c/a-1)}\left(\frac{b}{a}x\right) \tag{14}$$

With $\Gamma(\cdot)$ the gamma function and $L_n(x)$ the Laguerre polynomial defined as

$$L_n^{\alpha}(\alpha)(x) = \frac{(-1)^k \prod_{j=0}^{n-1}(\alpha+j+1)}{k!(n-k)! \prod_{j=0}^{k-1}(\alpha+j+1)} x^k$$

Our next step in solving this system, is finding a, b, c such that the entropy is maximal and all the boundary conditions are fulfilled.

Since t only occurs in the exponent in g we know for $s > t$ we have:

$$g(x,s \mid x(0) = x_0) \leq g(x,t \mid x(0) = x_0)$$

Since $\log g$ is an increasing function for $g \in (0,1]$ we know:t

$$\log g(x,s \mid x(0) = x_0) \leq \log g(x,t \mid x(0) = x_0)$$

Therefore, we also know:

$$-g(x,s \mid x(0) = x_0) \log g(x,s \mid x(0) = x_0) \geq -g(x,t \mid x(0) = x_0) \log g(x,t \mid x(0) = x_0)$$

So also

$$H(g(x,t \mid x(0) = x_0)) \leq H(g(x,s \mid x(0) = x_0))$$

This means that the entropy of g increases in t.

Since the entropy increases in t, we know that the maximum entropy probability density of $x(t)$ is given by the limit $t \to \infty$ of expression (14).
If we look at the at the limit $t \to \infty$ we see that all the terms in the summation become zero except for the $n = 0$ term. This leaves us with

$$\lim_{t\to\infty} g(x,t \mid x_0, 0) = \frac{b}{a}^{c/a} x^{c/a-1} e^{-\frac{b}{a}x} \frac{\Gamma(1)}{\Gamma(c/a)} L_0^{(c/a-1)}\left(\frac{b}{a}x_0\right) L_0^{(c/a-1)}\left(\frac{b}{a}x\right) \tag{15}$$

$$= \frac{b}{a}^{c/a} x^{c/a-1} e^{-\frac{b}{a}x} \frac{1}{\Gamma(c/a)} \tag{16}$$

$$= \gamma(x,t \mid c/a, b/a) \tag{17}$$

This is exactly the probability density function of a gamma distribution with parameters c/a and b/a.

As we have argued above, we know that the probability density of $S(T)$ is the same as that of $x(T)$. This leaves us with the maximum entropy probability distribution $\gamma(S(T)\ |u = c/a, v = b/a)$. To solve our system of equations, we still have to fulfill the boundary conditions for a, b, c. We calculate

$$\begin{aligned} \mathbb{E}(S(T)) &= \frac{v}{u} = S(0)/P(0) \\ Var(S(T)) &= \frac{v}{u^2} = \sigma^2 \end{aligned}$$

This leaves us with the following expressions for u and v

$$\begin{aligned} u &= \frac{(S(0)/P(0))}{\sigma^2} > 0 \\ v &= \frac{(S(0)/P(0))^2}{\sigma^2} > 0 \end{aligned}$$

Since u and v are both bigger then zero, this gamma density is well-defined. This completes our proof. □

3.4 Call and Put options

In this section we will derive the price of a put and call option under the risk-neutral probability density we have derived in the previous section. We will do this by a straightforward calculation. This section only contains the calculations. In the next chapter we will use these expressions to calculate the prices of options with real data.

Proposition 3.4.1. *The prices of European call and put options on dividend protected stocks with strike price K are*

$$\begin{aligned} Call &= S(0)\,(1 - G(K \mid u, v+1)) - P(0)K\,(1 - G(K \mid u, v)) \\ Put &= P(0)KG(K \mid u, v) - S(0)G(K \mid u, v+1) \end{aligned}$$

with u and v defined as in Proposition 3.3.4 and G the cumulative gamma distribution function

$$G(y \mid u, v) = \int_0^y \gamma(z \mid u, v)\, dz$$

Proof. We will calculate this. Recall that

$$u = \frac{(S/P)}{\sigma^2} \text{ and } v = \frac{(S/P)^2}{\sigma^2}$$

We calculate

$$\begin{aligned}
Call &= \mathbb{E}_\gamma[P(0)\max\{0, S(T)-K\}] \\
&= P(0)\int_{-\infty}^{\infty} \gamma(S(T)\mid u,v)\max\{0, S(T)-K\}\, dS(T) \\
&= P(0)\int_{-\infty}^{\infty} \gamma(S(T)\mid u,v)\mathbb{1}_{S(T)>K}(S(T)-K)\, dS(T) \\
&= P(0)\int_{K}^{\infty} \gamma(S(T)\mid u,v)(S(T)-K)\, dS(T) \\
&= P(0)\int_{K}^{\infty} \gamma(S(T)\mid u,v)S(T)\, dS(T) - P(0)K\,(1-G(K\mid u,v)) \\
&= P(0)\int_{K}^{\infty} \frac{u^v}{\Gamma(v)}S(T)^{v-1}e^{-uS(T)}S(T)\, dS(T) - P(0)K\,(1-G(K\mid u,v)) \\
&= P(0)\int_{K}^{\infty} \frac{v}{u}\frac{u^{v+1}}{\Gamma(v+1)}S(T)^{v}e^{-uS(T)}\, dS(T) - P(0)K\,(1-G(K\mid u,v)) \\
&= S(0)\cdot(1-G(K\mid u,v+1)) - P(0)\cdot K\cdot(1-G(K\mid u,v))
\end{aligned}$$

To compute the price of a put option we calculate

$$\begin{aligned}
Put &= \mathbb{E}_\gamma[P(0)\max\{0, K-S(T)\}] \\
&= P(0)\int_{0}^{K} \gamma(S(T)\mid u,v)(K-S(T))\, dS(T) \\
&= P(0)KG(K\mid u,v) - P(0)\int_{0}^{K} \gamma(S(T)\mid u,v)S(T)\, dS(T) \\
&= P(0)KG(K\mid u,v) - P(0)\frac{v}{u}G(K\mid u,v+1) \\
&= P(0)\cdot K\cdot G(K\mid u,v) - S(0)\cdot G(K\mid u,v+1)
\end{aligned}$$

This completes the proof. □

We will use these expressions in the next chapter.

In the next chapter we will numerically calculate Black-Scholes prices. For convenient purposes, we will state the Black-Scholes expressions for European call and put options. A derivation of these expressions can be found in *The Concepts and Practice of Mathematical Finance*[8], page 119-121.

Proposition 3.4.2. *Assume we have full knowledge of the information set* $\{S(0), P(0), \sigma^2\}$. *The Black-Scholes prices of European call and put options on dividend protected stocks with strike price K are*

$$\begin{aligned}
Call &= S(0)\cdot(1-N(d_1)) - P(0)\cdot K\cdot(1-N(d_2)) \\
Put &= P(0)\cdot K\cdot N(-d_2) - S(0)N(-d_1)
\end{aligned}$$

With d_1 and d_2 defined as:

$$d_1 = \frac{1}{\sigma\sqrt{T}}\Big(\log(S(0)/K) + (r + \sigma^2/2)T\Big)$$
$$d_2 = \frac{1}{\sigma\sqrt{T}}\Big(\log(S(0)/K) + (r - \sigma^2/2)T\Big)$$

and N the cumulative standard normal distribution.

4 Numerical research

In this chapter we will use proposition 3.4.1 from chapter 3 to make some numerical calculations. To asses whether or not our model is good we will compare it with the Black-Scholes model which is widely regarded as the most common model for option valuation.

We will do our numerical research in the Python programming environment. We will evaluate a call option at 60 dollars and a put option at 80 dollars.

We will evaluate one month of data. In our evaluation we use the implied volatility and the interest rate found on *https://countryeconomy.com/key-rates/usa*.

4.1 Call option 60 dollar

To calculate option prices, we need to calculate the variance of the underlying asset. We will estimate this variance by the formula:

$$\frac{1}{T}\sum_{j=0}^{m-1}\left(\log\frac{S(t_{j+1})}{S(t_j)}\right)^2 \approx \sigma^2$$

The derivation of this formula can be found in *Stochastic Calculus for Finance II*[12], page 107.

The standard way for evaluating option values in finance is by using the Black-Scholes model. If we compare this model with our entropy model we see in figure 1 that they are very similar. If we look very close at figure 1 we can see the graph of the entropy model. This means that our model matches the Black-Scholes model. This is a good sign since the Black-Scholes model is still a good indication for option pricing.

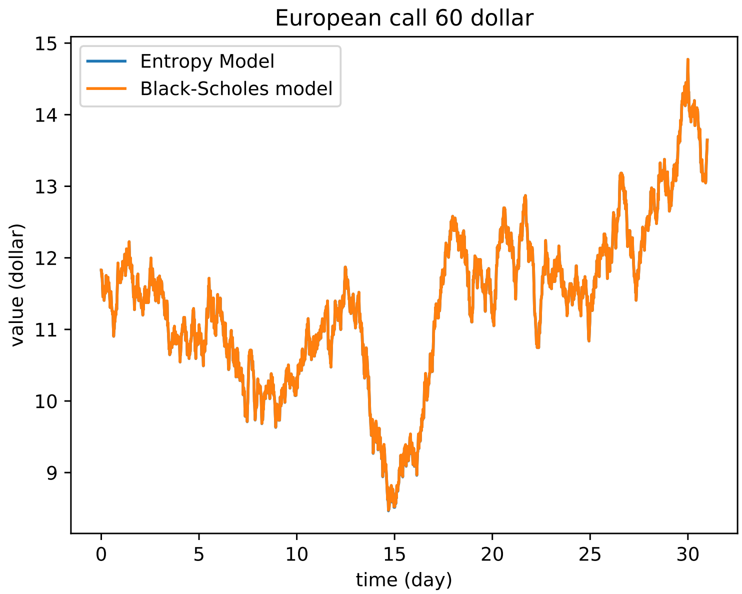

Figure 1: Price of an European call option

We also want to compare our model with the market prices of the option. This is interesting because if the entropy model price is different then the market price and we trust our model then it is easy to make profit by selling or buying the option.

If we look at figure 2 we see that the entropy model prices have the same behaviour but are constantly estimated lower then the real market prices. This means that by selling this call option we can make a profit worth the difference between both estimations.

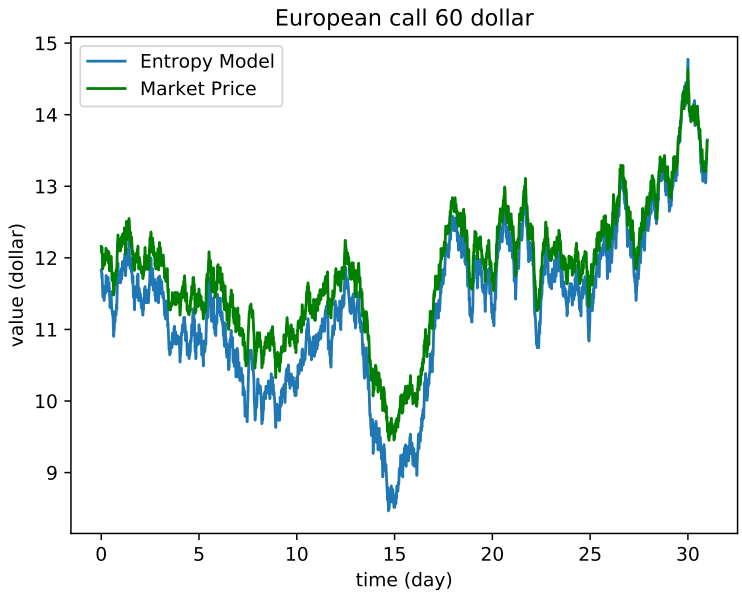

Figure 2: Price comparison of an European call option

4.2 Put option 80 dollar

In this section we will look at a put option with a strike price of 80 dollar. We will calculate the option prices and compare them with the option prices that are provided by the Black-Scholes model and the market prices.

If we look at figure 3 we see that the prices calculated with the entropy model and the prices calculated with the Black-Scholes model look very much the same. Just as with the call option we see that both graphes show the same behaviour.

If we look at the comparison of the entropy prices with the market prices in figure 4 we see that both prices have the same behaviour. The entropy prices are a lot lower then the Black-Scholes prices. We have also seen this at the call option of 60 dollar.

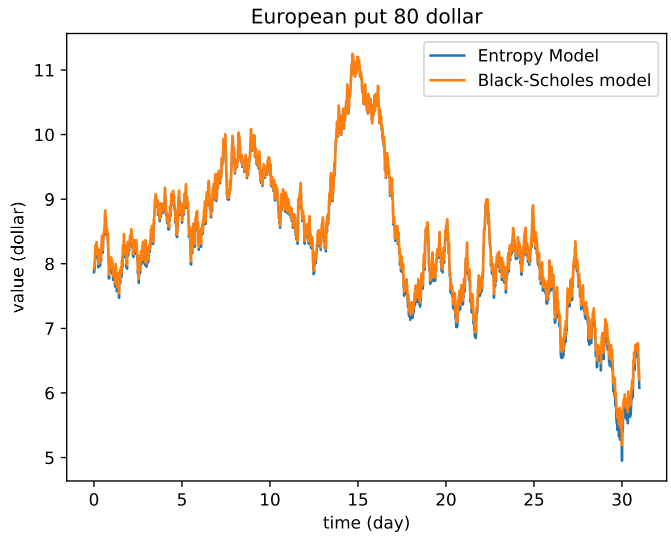

Figure 3: Price of an European put option

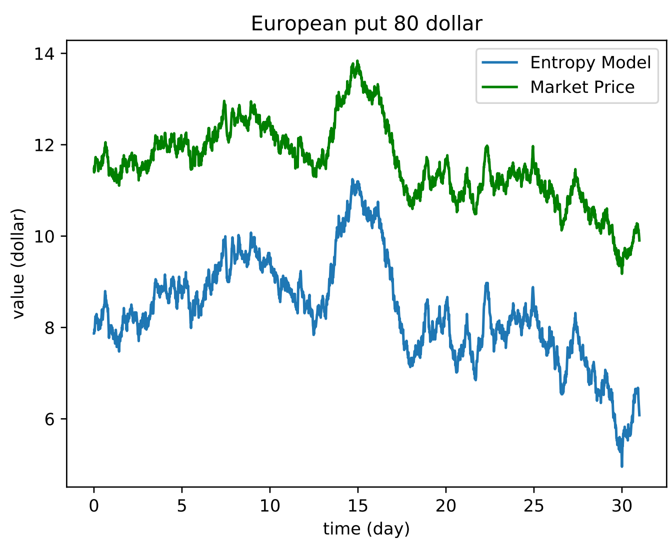

Figure 4: Price comparison of an European put option

4.3 Discussion

The reason for the difference between both prices could be that we have determined our parameters wrong. A higher interest rate for example would provide higher entropy prices and higher Black-Scholes prices.

A reason for the difference in the prices at the put option could be that the market is always afraid for losses. Therefore players in the market will always asses the risk of loss higher then it actually is. This results in a higher market price for the put option. One downside of this explanation is that at the call option the market price is still higher then the entropy price. If one would asses the risk of loss higher then the risk of profit we would expect that the market price would be lower then the entropy price.

5 Conclusion

In chapter 1 of this thesis we have given a brief introduction to measure theory. After that we have gone further and introduced the notion of a dynamical system. With the discussion of ergodic theory we have seen that there exists special measures named ergodic measures. After that we have proven that under certain conditions an ergodic measure always exists.

With the theory from chapter 1 we have defined what entropy is. We have seen that entropy is a way to measure disorder in a dynamical system. Later we have proved that we can use entropy to distinguish dynamical systems from one another. If two dynamical systems are isomorphic they have the same entropy.

We have used entropy to determine the prices of call and put options on the financial markets. We have seen that by solving a system of equations we can derive the risk neutral probability of a stock. Instead of obtaining the lognormal density we get in the Black-Scholes model we have obtained a gamma probability density.

In the end of this thesis we have done some numerical research to see if our model works with real stock data. We have evaluated option prices with our entropy model and compared them with the prices we get with the Black-Scholes model and the real market prices. We have seen that the prices from the entropy model are very similar to the prices we get from the Black-Scholes model. This means that our model is not better then the Black-Scholes model and that we therefore can not conclude that we can replace the Black-Scholes model with the entropy model. We have also seen that both our entropy model and the Black-Scholes model constantly evaluate option prices lower then the market prices. If this is true one can make profit by selling options for the market price because this price would be higher then the fair price. To see if this is really true one needs to do more numerical research. For example one can evaluate more options. Also one can simulate the hedging proces one can derive from the entropy model.

We conclude that we have find an alternative model of the Black-Scholes model. This model gives us the same prices as the Black-Scholes model and therefore it is not better then the Black-Scholes model. This model does give us more confidence in the Black-Scholes prices since we have now derived these prices in two different ways.

A Appendix

A.1 Appendix Chapter 1

Theorem A.1.1 (Regularity of $M(X,\mathcal{B})$)**.** *Every member of $M(X,\mathcal{B})$ is regular, i.e. for all $B \in \mathcal{B}$ and every $\epsilon > 0$ there exist an open set U_ϵ and a closed set C_ϵ such that $C_\epsilon \subseteq B \subseteq U_\epsilon$ such that $\mu(U_\epsilon \backslash C_\epsilon) < \epsilon$.*

Proof. We call a set with the above property a regular set. Let $\mathcal{R} = \{B \in \mathcal{B} \mid B \text{is regular}\}$. We will first show that $\mathcal{R}$ is a σ-algebra. After that we will show that $\mathcal{R}$ contains all closed sets.

Obviously $\emptyset \in \mathcal{R}$. Now take $A \in \mathcal{R}$ and let $\epsilon > 0$. Take C closed and U open with $C \subset A \subseteq U$ and

$$\mu(A) < \mu(C) + \epsilon$$
$$\mu(A) > \mu(U) - \epsilon$$

We see that $U^c \subset A^c \subset U^c$. By definition we know that U^c is closed and C^c is open. Also

$$\begin{aligned}\mu(A^c) = \mu(X) - \mu(A) &< \mu(X) - \mu(U) + \epsilon \\ &= \mu(C^c) + \epsilon \\ \mu(A^c) = \mu(X) - \mu(A) &> \mu(X) - \mu(U) \\ &= \mu(U^c) - \epsilon\end{aligned}$$

So $A^c \in \mathcal{R}$

Take $A_1, A_2, ... \in \mathcal{R}$ and let $\epsilon > 0$. Take for each $i \in \mathbb{N}$ a U_i open and a C_i closed such that $C_i \subset A \subset U_i$ and

$$\mu(U_i) - \mu(A_i) < 2^{-i}\epsilon$$
$$\mu(A_i) - \mu(C_i) < 2^{-i}\epsilon/2$$

We know this is possible since every $A_i \in \mathcal{R}$.

We see that $\cup_i C_i \subset \cup_i A_i \subset \cup_i U_i$ and $\cup_i U_i$ is open. We calculate

$$\begin{aligned}\mu(\cup_i U_i) - \mu(\cup_i A_i) &\leq \mu(\cup_{i=1}^\infty U_i \backslash \cup_{i=1}^\infty A_i) \\ &\leq \mu(\cup_{i=1}^\infty (U_i \backslash A_i)) \\ &\leq \sum_{i=1}^\infty \mu(U_i \backslash A_i) \\ &= \sum_{i=1}^\infty (\mu(U_i) - \mu(A_i)) \\ &< \sum_{i=1}^\infty 2^{-i}\epsilon = \epsilon\end{aligned}$$

Since a measure is continuous we know that $\lim_{k\to\infty} \mu(\cup_{i=1}^{\infty} C_i) = \mu(\cup_{i=1}^{\infty})$. So for some large $k \in \mathbb{N}$ we have $\mu(\cup_{i=1}^{\infty} C_i) - \mu(\cup_{i=1}^{k} C_i) < \epsilon/2$. We see that $\cup_{i=1}^{k} C_i$ is a finite sum of closed sets, so $\cup_{i=1}^{\infty} C_i$ is closed. We calculate

$$\begin{aligned}\mu(\cup_{i=1}^{\infty} A_i) - \mu(\cup_{i=1}^{k} C_i) &< \mu(\cup_{i=1}^{\infty} A_i) - \mu(\cup_{i=1}^{\infty} C_i) + \epsilon/2 \\ &\leq \mu(\cup_{i=1}^{\infty}(A_i \backslash C_i)) + \epsilon/2 \\ &\leq \sum_{i=1}^{\infty}(\mu(A_i) - \mu(C_i)) + \epsilon/2 \\ &\leq \epsilon/2 + \epsilon/2 = \epsilon\end{aligned}$$

Hence $\cup_{i=1}^{\infty} \in \mathcal{R}$. So $\mathcal{R}$ is a σ-algebra.

Let $A \subset X$ be closed. Let $U_n = \{x \in X \mid \exists a \in A \text{ with } d(a,x) < \frac{1}{n}\}$. We see that U_n is open, $U_1 \supset U_2 \supset ...$ and $\cap_{i=1}^{\infty} U_i = A$ because A is closed. Hence we have $\mu(A) = \lim_{n\to\infty} \mu(U_n) = \inf_i \mu(U_n)$. So

$$\begin{aligned}\mu(A) &\leq \inf\{mu(U) \mid U \supset A,\ U\text{open}\} \\ &\leq \inf_n \mu(U_n) = \mu(A)\end{aligned}$$

So $A \in \mathcal{R}$

Since $\mathcal{R}$ is a σ-algebra that contains all closed sets we know that $\mathcal{B} \subset \mathcal{R}$. □

Corollary A.1.2. *For any $B \in \mathcal{B}$, and any $\mu \in M(X, \mathcal{B})$ we have*

$$\mu(B) = \sup_{C \subseteq B \mid C \text{ closed}} \mu(C) = \inf_{B \subseteq U \mid U \text{ open}}$$

A.2 Appendix Chapter 2

Proposition A.2.1. *[Proposition A.2.1] Let $(X, \mathcal{A}, \mu, T)$ be a measure preserving dynamical system. Let $\alpha, \beta, \gamma \in \Omega_X$. Then:*

(i) $H(T^{-1}\alpha) = H(\alpha)$

(ii) $H(\alpha \vee \beta) = H(\alpha) + H(\beta \mid \alpha)$

(iii) $H(\beta \mid \alpha) \leq H(\beta)$

(iv) $H(\alpha \vee \beta) \leq H(\alpha) + H(\beta)$

(v) If $\alpha \leq \beta$ *then* $H(\alpha) \leq H(\beta)$

(vi) $H(\alpha \vee \beta \mid \gamma) = H(\alpha \mid \gamma) + H(\beta \mid \alpha \vee \gamma)$

(vii) If $\beta \leq \alpha$ *then* $H(\gamma \mid \alpha) \leq H(\gamma \mid \beta)$

(viii) If $\beta \leq \alpha$ *then* $H(\beta \mid \alpha) = 0$

Proof. (i) $H(T^{-1}\alpha) = -\sum_{A\in\alpha} \mu(T^{-1}A)\log\mu(T^{-1}A)$. But since T is measure preserving we know that $H(T^{-1}\alpha) = -\sum_{A\in\alpha} \mu(A)\log\mu(A) = H(\alpha)$

(ii)

$$\begin{aligned}
H(\alpha \vee \beta) &= -\sum_{A\in\alpha}\sum_{B\in\beta} \mu(A\cap B)\log\mu(A\cap B) \\
&= -\sum_{A\in\alpha}\sum_{B\in\beta} \mu(A\cap B)\log\left(\frac{\mu(A\cap B)}{\mu(A)}\right) - \sum_{A\in\alpha}\sum_{B\in\beta} \mu(A\cap B)\log\mu(A) \\
&= H(\beta \mid \alpha) + H(\alpha)
\end{aligned}$$

(iii) First we define $g(t) = -t\log(t)$. We note that g is concave for $0 < t \leq 1$. Now we calculate:

$$\begin{aligned}
H(\beta \mid \alpha) &= -\sum_{A\in\alpha}\sum_{B\in\beta} \mu(A\cap B)\log\left(\frac{\mu(A\cap B)}{\mu(A)}\right) \\
&= -\sum_{A\in\alpha}\sum_{B\in\beta} \mu(A)\frac{\mu(A\cap B)}{\mu(A)}\log\left(\frac{\mu(A\cap B)}{\mu(A)}\right) \\
&= \sum_{A\in\alpha}\sum_{B\in\beta} \mu(A) g\left(\frac{\mu(A\cap B)}{\mu(A)}\right)
\end{aligned}$$

Since g is concave for $t = \frac{\mu(A\cap B)}{\mu(A)}$ we know that we can approximate this as:

$$\begin{aligned}
&\leq \sum_{B\in\beta} g\left(\sum_{A\in\alpha} \mu(A)\frac{\mu(A\cap B)}{\mu(A)}\right) \\
&\leq \sum_{B\in\beta} g\left(\sum_{A\in\alpha} \mu(A\cap B)\right) \\
&= H(\mathcal{B})
\end{aligned}$$

(iv) This result is trivial if we use ii. and iii.

(v) First observe that if $\alpha \leq \beta$ then $\beta = \alpha \vee \beta$. Therefore we can calculate

$$H(\beta) = H(\alpha \vee \beta) = H(\alpha) + H(\beta \mid \alpha) \geq H(\alpha)$$

(vi) again since $\beta \leq \alpha$ we have that $\alpha = \alpha \vee \beta$. So we can calculate

$$\begin{aligned}
H(\gamma \mid \alpha) &= H(\gamma \mid \alpha \vee \beta) \\
&= H(\alpha \vee \gamma \mid \beta) - H(\alpha \mid \beta) \\
&\leq H(\alpha \mid \beta) + h(\gamma \mid \beta) - H(\alpha \mid \beta) \\
&= H(\gamma \mid \beta)
\end{aligned}$$

(vii) Again we have $\alpha = \alpha \vee \beta$. We calculate

$$\begin{aligned}
H(\beta \mid \alpha) &= H(\beta \mid \alpha \vee \beta) \\
&= H(\alpha \vee \beta \mid \beta) - H(\alpha \mid \beta) \\
&\leq H(\alpha \mid \beta) + H(\beta \mid \beta) - H(\alpha \mid \beta) = 0
\end{aligned}$$

□

A.3 Python code Chapter 4

```
import os
import csv
import numpy as np
import pandas as pd
from scipy import stats
import matplotlib.pyplot as plt

os.chdir("/home/niek/PycharmProjects/HelloWorld")
csv_file2 = open('data.csv')
csv_reader2 = csv.reader(csv_file2, delimiter=',')

data = [xi for xi in csv_reader2]
del (data[0])
data1 = data[0:8927:3]
data2 = data[8927:-1:3]

# Making differet lists of our data
date = np.array([i[0] for i in data1])
expirationtime = np.array([float(i[5]) for i in data1])
stockprices1 = np.array([(float(i[3]) + float(i[1])) / 2. for i in data1])
stockprices2 = np.array([(float(i[3]) + float(i[1])) / 2. for i in data2])
calloption60price = np.array([(float(i[18]) + float(i[20])) / 2. for i in data1])
putoption60price = np.array([(float(i[6]) + float(i[8])) / 2. for i in data1])
calloption70price = np.array([(float(i[22]) + float(i[24])) / 2. for i in data1])
putoption70price = np.array([(float(i[10]) + float(i[12])) / 2. for i in data1])
calloption80price = np.array([(float(i[26]) + float(i[28])) / 2. for i in data1])
putoption80price = np.array([(float(i[14]) + float(i[16])) / 2. for i in data1])


# A function to calculate our estimated variance
def sigma_calculator(lst, T):
    lst2 = []
    for i in range(0, len(lst) - 1):
        lst2.append(np.log(float(lst[i + 1]) / float(lst[i])) ** 2)
    sigma = np.sqrt(1. / T * np.sum(lst2))
    return sigma


# A function that calculates option prices under our entropy model
def entropy_price_calculator(lst, type):
    S = float(lst[0])
    r = lst[1]
    sigma = lst[2]
    K = lst[3]
    T = lst[4]

    P = np.exp(-r * T)
    u = (S / P) / (sigma ** 2)
```

```
    v = ((S / P) ** 2) / (sigma ** 2)
    if type == 'call':
        return S * (1. - stats.gamma.cdf(K, v + 1, scale=1. / float(u))) - P * K * (
                        1. - stats.gamma.cdf(K, v, scale=1. / float(u)))
    if type == 'put':
        return P * K * stats.gamma.cdf(K, v, scale=1. / float(u)) - S * stats.gamma.c


# A function that calculates option prices under the Black and Scholes model
def black_and_scholes_calculator(lst, type):
    S = float(lst[0])
    r = lst[1]
    sigma = lst[2]
    K = lst[3]
    T = lst[4]

    d1 = (np.log(S / K) + (r + 0.5 * (sigma ** 2.0)) * T) / (sigma * np.sqrt(T))
    d2 = (np.log(S / K) + (r - 0.5 * (sigma ** 2.0)) * T) / (sigma * np.sqrt(T))
    if type == 'call':
        return S * stats.norm.cdf(d1) - K * np.exp(-r * T) * stats.norm.cdf(d2)
    if type == 'put':
        return K * np.exp(-r * T) * stats.norm.cdf(-d2) - S * stats.norm.cdf(-d1)


datalst = []
for i in range(0, len(data1)):
    lst = [
        stockprices1[i],
        0.025 / 12. * 8.,
        sigma_calculator(stockprices2, 1),
        60.,
        float(expirationtime[i])
    ]
    datalst.append(lst)

# Call option
lst_black_and_scholes = datalst[:]
lst_black_and_scholes_prices_call = np.array([black_and_scholes_calculator(xi, 'call'
lst_black_and_scholes_prices_call = lst_black_and_scholes_prices_call[1:-1]

lst_entropy = datalst[:]
lst_entropy_prices_call = np.array([entropy_price_calculator(xi, 'call') for xi in ls
lst_entropy_prices_call = lst_entropy_prices_call[1:-1]
calloption60price = calloption60price[1:-1]

# Put option.
lst_black_and_scholes_put = datalst[:]
for i in range(1, len(lst_black_and_scholes_put)):
    lst_black_and_scholes_put[i][3] = 80.
lst_black_and_scholes_prices_put = np.array(
```

```
    [black_and_scholes_calculator(xi, 'put') for xi in lst_black_and_scholes_put])
lst_black_and_scholes_prices_put = lst_black_and_scholes_prices_put[1:-1]

lst_entropy_put = datalst[:]
for i in range(1, len(lst_entropy_put)):
    lst_entropy_put[i][3] = 80
lst_entropy_prices_put = np.array([entropy_price_calculator(xi, 'put') for xi in lst_
lst_entropy_prices_put = lst_entropy_prices_put[1:-1]
putoption80price = putoption80price[1:-1]

t = np.arange(0., 31., 31. / float(len(lst_entropy_prices_call)))

plt.plot(t, lst_entropy_prices_call)
plt.plot(t, lst_black_and_scholes_prices_call)
plt.gca().legend(('Entropy Model', 'Black-Scholes model'))
plt.title('European call 60 dollar')
plt.xlabel('time (day)')
plt.ylabel('value (dollar)')
plt.show()

plt.plot(t, lst_entropy_prices_call)
plt.plot(t, calloption60price, color='green')
plt.gca().legend(('Entropy Model', 'Market Price '))
ax = plt.gca()
leg = ax.get_legend()
leg.legendHandles[1].set_color('green')
plt.title('European call 60 dollar')
plt.xlabel('time (day)')
plt.ylabel('value (dollar)')
plt.show()

plt.plot(t, lst_entropy_prices_put)
plt.plot(t, lst_black_and_scholes_prices_put)
plt.gca().legend(('Entropy Model', 'Black-Scholes model'))
plt.title('European put 80 dollar')
plt.xlabel('time (day)')
plt.ylabel('value (dollar)')
plt.show()

plt.plot(t, lst_entropy_prices_put)
plt.plot(t, putoption80price, color='green')
plt.gca().legend(('Entropy Model', 'Market Price '))
ax2 = plt.gca()
leg = ax2.get_legend()
leg.legendHandles[1].set_color('green')
plt.title('European put 80 dollar')
plt.xlabel('time (day)')
plt.ylabel('value (dollar)')
plt.show()
```

References

[1] Barney Bramham. *Notes for the Course: Ergodic Theory and Entropy*. url: https://www.ruhr-uni-bochum.de/imperia/md/content/mathematik/Bramham/ergodic_notes.pdf. Bochum, Germany: Ruhr-Universität Bochum, 2014.

[2] Donald L. Cohn. *Measure Theory*. Suffolk University: Birkhauser Boston, 1980. ISBN: 3-7643-3003-1.

[3] John B. Conway. *A course in Functional Analysis*. 2nd ed. Springer, 1990. ISBN: 0-387-97245-5.

[4] Karma Dajani and Sjoerd Dirksen. *Introduction to Ergodic Theory*. 1st ed. url: http://www.staff.science.uu.nl/~kraai101/lecturenotes2009.pdf. Utrecht University, Dec. 2018.

[5] Robert M. Gray. *Entropy and Information Theory*. 1st ed. url: https://ee.stanford.edu/~gray/it.pdf. Stanford University, Electrical Engineering Department: Springer-Verlag, 2014.

[6] Les Gulko. "The Entropy Theory of Stock Option Pricing". In: *Yale School of Management* (Oct. 1997).

[7] Chris Hillman. "An Entropy Primer". In: (June 1996). url: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.54.517&rep=rep1&type=pdf.

[8] Mark S. Joshi. *The Concepts and Practice of Mathematical Finance*. Vol. 2. Cambridge: Cambridge University Press, 2008. ISBN: 0 521 82355 2.

[9] Samuel Karlin and Howard M. Taylor. *A Second Course in Stochastic Processes*. Academic Press, 1981. ISBN: 0-12-398650-8.

[10] David G. Luenberger. *Optimization by Vector Space Methods*. New York: John Wiley, 1969. ISBN: 68-8716 SBN 471 55359x.

[11] Sheldon M. Ross. *Stochastic Processes*. University Of California: John Wiley & Sons, 1983. ISBN: 0-471-09942-2.

[12] Steven E. Shreve. *Stochastic Calculus for Finance 2*. New York: Springer, 19692004. ISBN: 978-0-387-40101-0.

[13] V.S. Sunder. "The Riesz representation theorem". In: *Institute of Mathematical Sciences, Madras, India* (Apr. 2008). url: https://www.imsc.res.in/~sunder/rrt1.pdf.