Bayesian Learning in Mis-specified Models

Maarten-Pieter Schinkel * Jan Tuinstra † Dries Vermeulen ‡

February 21, 2000

preliminary version

Abstract

A central unanswered question in economic theory is that of price formation in disequilibrium. This paper lays down the methodological groundwork for a model that has been suggested as an answer to this question (Arrow, 1959; Fisher, 1983; Hahn, 1989). We consider sellers that monopolistically compete in prices but have incomplete information about the structure of the market they face. They each entertain a simple demand conjecture in which sales are perceived to depend on the own price only, and set prices to maximize expected profits. Prior beliefs on the parameters of conjectured demand are updated into posterior beliefs upon each observation of sales at proposed prices, using Bayes' rule. The rational learning process thus constructed drives the price dynamics of the model. Its properties are analysed. Moreover, a sufficient condition is provided, relating objectively possible events and subjective beliefs, under which the price process is globally stable on a conjectural equilibrium for almost all objectively possible developments of history.

Journal of Economic Literature Classification Numbers: C11, C51, C62, D21, D43, D83.

KEYWORDS: Oligopolistic price setting, Bayesian learning, Conjectural equilibrium, Global Stability.

1. Introduction

In economic theory, a key role in the coordination of behaviour is played by prices. As a consequence, the so-called price mechanism is much debated, and

*University of Maastricht, Dept. of Economics, e-mail: m.schinkel@algec.unimaas.nl.

†University of Amsterdam, Institute of Actuarial Science and Econometrics, e-mail: tuinstra@fee.uva.nl.

‡University of Maastricht, Dept. of Economics, e-mail: d.vermeulen@algec.unimaas.nl.

the need for it operating freely often stressed. Yet there are many open research questions on the matter of prices, especially on how they come to take on equilibrium values. For one thing, it is generally left unexplained whose business it actually is to call and change prices. Particularly in models in which price-taking behaviour is assumed, this is a pressing question. Reliance on a unique price vector indicates it is left to a single person or institution, and a number of models has been presented in which the central person is in fact an altruistic auctioneer—*e.g.* in the tâtonnement process, the Edgeworth process, and the Hahn process.

Apart from the fact that it seems odd, if not plainly inconsistent, to model all behaviour but that of the auctioneer as resulting from constrained rational choice, at least two things meet the eye in these explanations. First, they need an exogenous central coordinator to explain the rise of equilibria that are meant to be the outcome of decentralized competitive economies. Second, the conditions these processes need for convergence on equilibrium price values for arbitrary initial prices—*i.e.* for global stability of the disequilibrium process—have been found to be pretty strong.

A number of suggestions has been made to study the disequilibrium behaviour of prices more seriously. An early one was by Arrow (1959). He proposed to make price a choice variable of individual firms, that consequently need to come equipped with some local monopoly power, at least as a disequilibrium phenomenon. To Arrow, the construct of perfect competition did not allow for an explanation of price behaviour. More recently, Fisher (1983) developed an elaborate model of disequilibrium behaviour in which there is clarity on who is setting prices. It is done by dealers, who specialize in differentiated goods, which gives them the local monopoly to act as a coordinator. Fisher's objective to have disequilibrium processes end in competitive equilibrium, however, led him originally to model dealers as little auctioneers, changing prices in the direction of competitive equilibrium values. Yet, the general structure of his stability proof allows for dealers to set prices more rationally, exploiting their powers, and this is done later in the book. How prices change with perceptions, however, is not discussed in depth. In Hahn (1989) several partial examples are given of perception changes and associated behaviour that may indeed be plausible for monopolistically competing price setters to develop—including a rudimentary version of the behaviour we study in this paper. Yet, the consequences of such behaviour, particularly when performed in general equilibrium settings, are only hinted upon.

When prices are choice variables of firms, the way firms perceive their market position, and especially changes in these perceptions, can account for the dynamics of prices. This idea is used in the present paper to construct a model of price adjustment and study its limit behaviour, *i.e.* its stability properties. In the present model, a number of firms is in monopolistic price competition, but does not have perfect information on the market demand it faces. At each moment in time, based on its information to date on past prices and sales, each firm entertains a demand conjecture instead. Naturally, this conjecture has a structural form different from that of objective demand. Particularly, we consider the most extreme case where firms only consider their own price as an explanatory variable, and do not consider the price effects of competing products. Within their conjectured structures, firms learn in a Bayesian way about the value of the demand parameters it has modelled. A fleshed out conjecture serves as a basis for an optimal price through expected profit maximization.

It is shown that, for initial beliefs that do not assign zero probability to developments of prices and sales that can actually happen, the incomplete beliefs converge to a finite limit, and therefore prices converge as well. This is called 'no statistical surprise'. Convergence takes place on a set of 'conjectural equilibria'. Under 'no statistical surprise', therefore, the price process is globally stable in that it reaches an equilibrium for every initial belief-structure. Which particular equilibrium is reached depends on the initial beliefs. This path-dependency result runs solely over beliefs, since the model assumes the absence of trade at disequilibrium prices. The stability result does not rely on specific conditions on the structure of objective demand. Instead, the 'no statistical surprise' condition is sufficient for the perceived structure to absorb all price effects on objective demand.

The literature on Bayesian or rational learning is quite recent and large. Our paper builds on several of its results. One focus has been the concern to justify the use of rational expectations equilibria. Particularly Bray and Savin (1986), and Bray and Kreps (1987) have worked in this direction, and have established convergence results for myopic Bayesian learners on rational expectations equilibrium in versions of the cobweb-model. Early work by Blume and Easley (1982; 1984) is also concerned with the influence learning has on the eventual equilibrium situation reached, but in a general equilibrium setting. Particularly, they have focussed on conditions under which Bayesian learners will identify the true model among several models.

In partial equilibrium models of single firms learning their demand, Easley and Kiefer (1988) among others, study the influence of active learning on firms' optimization problems. Actively learning firms are aware of the fact that their behaviour influences their options for learning. In a discrete game theoretical setting, Kalai and Lehrer (1993; 1995) have obtained results for rational learning behaviour. Kalai and Lehrer (1993) considers learning in a correctly specified structure, and states conditions under which it converges to a Nash equilibrium of the perfect information game that are similar to ours.

Another, much less extensively travelled, route has been to study the influence of structural misspecification on the convergence process and its equilibria. Kirman (1975; 1983; 1995) sets up an early example of two firms learning, in a least squares way, in a misspecified structure of their game. He does not establish general convergence results, however. Nyarko (1991) constructs an example of a single, actively learning monopolist whose beliefs do not settle, due to a very particular structural specification error. Kalai and Lehrer (1995) extends the 1993 convergence conditions to structurally misspecified models to identify the usable notion of equilibrium. The article does not present explicit convergence results, however.

This paper is organized as follows. The next section presents the model structure. Section 3 discusses some elementary properties of stochastic processes. Sections 4 and 5 present the convergence result, and discuss its nature. Section 6 concludes on the global stability of the price process on the equilibria of the model, introducing the concept of 'no statistical surprise'. Section 7 summarizes.

2. The Model

Consider an economy with n different firms. Each firm i has the ability to produce its own commodity. It is supposed to take decisions on price, quality, etc. concerning the commodity it produces. In this paper we will assume that the aggregate of all these strategic choices to be made by firm i are incorporated into one single action space P_i. For technical reasons each P_i is assumed to be a compact metric space. [1]

OBJECTIVE DEMAND

In this paper we will assume that the objective demand for commodity i is not deterministic. In order to model this, let the commodity space of firm i be denoted by X_i. For technical reasons this commodity space is assumed to be a compact metric space as well.

Suppose that firm i has decided to take action p_i in P_i. We write $p := (p_i)_{i \in N} \in P := \prod_i P_i$ for the entire vector of decisions taken. Now the demand for commodity i is supposed to be given by the density function

$$f_i(x_i \mid p)$$

with respect to the probability measure ν_i defined on the Borel σ-algebra $\mathcal{B}(X_i)$ generated by the metric on the commodity space X_i.

Remarks. For technical reasons we assume that for any open set $U \subset X_i$ we have $\nu_i(U) > 0$. Further, by $f_i(x_i \mid p)$ being defined with respect to ν_i we mean

[1] The paper applies a variety of concepts from real analysis. In order to make the paper self-contained, we offer them in an appendix.

that

$$\int_{X_i} f_i(x_i \mid p) d\nu_i = 1.$$

We will also assume that the function $f_i: X_i \times P \to \mathbb{R}$ is continuous. ◁

PERCEIVED DEMAND

None of the firms is fully aware of the mechanism that generates the demand it faces. Instead, each firm i has a collection Θ_i of "worlds" it deems possible. In world $\theta_i \in \Theta_i$ it conjectures that it serves a demand function that is distributed according to the density function

$$g_i(x_i \mid p_i, \theta_i)$$

with respect to ν_i. Again, we assume for technical reasons that Θ_i is a compact metric space and that $g_i: X_i \times P_i \times \Theta_i \to \mathbb{R}$ is continuous.

Remarks. Subjective demand conjectures deviate importantly from objective demand: each firm only considers the effect of its own decision on the demand for its commodity, and neglects the influence of the decisions of the other commodities. In effect, each firm believes that it is a monopolist on its own market.

This structural misspecification reflects incomplete information on the side of the firms. We focus on this extreme situation where only the effect of a firm's own decision is considered for reasons of exposition. The analysis could be extended to include less severe forms of incomplete information, *e.g.* structures in which the effects of the actions taken by several of the nearest competitors are included. ◁

EXPECTED PROFITS

Within its structural misspecification of how the world works, each firm i believes that there exists a "true" world. However it does not know which of possible worlds in Θ_i is the true one. Instead, the firm's perception of the world is stochastic. This means that each firm i has a belief represented by an element of the set $\mathbb{P}(\Theta_i)$ of probability measures on Θ_i. Such a belief $\mu_i \in \mathbb{P}(\Theta_i)$ assigns to each Borel subset A of Θ_i a real number $\mu_i(A)$ that reflects the probability firm i assigns to the event that the real world is an element of A.

Further, let

$$\pi_i(p_i, x_i) \in \mathbb{R}$$

be the net profit of demand x_i when firm i decides to take action p_i. (We will assume throughout the paper that π_i is continuous.) Then, given a belief μ_i of firm i, the amount $\Pi_i(p_i, \mu_i)$ of money firm i expects to earn is given by

$$\Pi_i(p_i, \mu_i) = \int_{\Theta_i} \int_{X_i} \pi_i(p_i, x_i) g_i(x_i, \theta_i \mid p_i) d\nu_i d\mu_i.$$

Since each firm i is assumed to be rational it will try to maximize $\Pi_i(p_i, \mu_i)$ and take an optimal decision. Concerning optimal decisions we make the following assumption.

Assumption 1. Given the belief μ_i of firm i there is a unique optimal decision. In other words, there is exactly one decision in P_i, denoted by $p_i(\mu_i)$, for which $\Pi_i(p_i(\mu_i), \mu_i)$ is larger than or equal to $\Pi(p_i, \mu_i)$ for any other possible action p_i of firm i in P_i.

Remarks. Note that $p_i(\mu_i)$ need not maximize expected profits in an objective sense. This is so since, although the world is in fact stochastic, it is stochastic in a way different from perception. More specifically, given the vector $p(\mu) := (p_i(\mu_i))_{i \in N}$ of individual decisions, objective demand is distributed on X_i according to

$$f_i(x_i \mid p(\mu)),$$

which shows how the true sales opportunities depend on the beliefs of all firms. And in turn these opportunities determine the objective expected net profit. In other words, the objective expected net profit of firm i is in fact given by

$$\int_{X_i} \pi_i(p_i, x_i) f_i(x_i \mid p(\mu)) d\nu_i.$$

No firm is, of course, capable of tuning its behaviour to this true expected net profit. $\triangleleft$

3. Information processing and the Bayes operator

Beliefs are updated according to the Bayesian updating rule, as follows. Suppose that μ_i is the current belief of firm i in $\mathbb{P}(\Theta_i)$. Now the observation of demand x_i in X_i induces the updated belief $B_i(\mu_i)(x_i)$ in $\mathbb{P}(\Theta_i)$ that assigns to a Borel set $A \subset \Theta_i$ the probability

$$B_i(\mu_i)(x_i)(A) := \frac{\int_A g_i(\theta_i \mid p_i(\mu_i), x_i) d\mu_i}{\int_{\Theta_i} g_i(\theta_i \mid p_i(\mu_i), x_i) d\mu_i}.$$

Provided of course that the denominator is not equal to zero. In order to guarantee that this is the case, independent of the belief μ_i, we make the following assumption.

Assumption 2. For all p_i, θ_i and x_i,

$$g_i(x_i \mid p_i, \theta_i) > 0. \qquad \triangleleft$$

Given this assumption it can be shown that the above formula indeed yields a mapping

$$B_i \colon \mathbb{P}(\Theta_i) \times X_i \to \mathbb{P}(\Theta_i),$$

from the space of probability measures times the space of quantities X_i back to the space of probability measures. [2] This particular updating method, known as Bayesian updating, is firmly founded in probability theory. In that sense it is sensible from the firms' perspective to extract information from past observations in this way.

CONJECTURAL EQUILIBRIUM

Although it makes perfect sense from the perspective of the firms, the learning process described is ill-founded in objective terms since it is based on an unrecognized structural misperception of demand. Hence, in general it cannot be hoped that subjective perceptions will come to explain the true demand for a commodity. Yet, there is a natural candidate for beliefs that are in 'equilibrium' with the objective world. Consider a single firm. The firm's beliefs are in

[2] The technicalities supporting this statement can be found in Appendix D.

equilibrium if perceived optimal decisions set on the basis of this belief return quantities that are no ground for a revision of beliefs. This is the concept of individual conjectural equilibrium.

Definition 1. An *individual conjectural equilibrium* for firm i is a belief μ_i for which for all $x_i \in X_i$

$$B_i(\mu_i)(x_i) = \mu_i. \qquad \triangleleft$$

Since the observed sales depend upon the decisions of all firms, it is quite special for a single firm to be in individual conjectural equilibrium. Yet, if all firms simultaneously are in individual conjectural equilibrium, none has a reason to deviate unilaterally from its decision, since none believes it can improve its position by doing so. This leads us to consider the following notion of an equilibrium for our economy.

Definition 2. A *conjectural equilibrium* is a vector $\mu = (\mu_i)_{i \in N}$ of individual conjectural equilibria. $\triangleleft$

4. Learning dynamics

In the previous section we saw that firms have a mis-specified model of the true state of the world and they are not aware of this false interpretation of their environment. Nevertheless, given their mis-specification of the way the world works, they are aware of the fact that they are not fully informed about the true state of the world. This lack of information is modeled as a probability distribution μ_{i0} (the initial belief) over the collection Θ_i of all worlds that firm i deems possible. This belief reflects the amount of prior information firm i has concerning the true state of world.

Now since each firm is a profit maximizer and since it is aware of the fact that it is not fully informed, it is eager to learn more about the true state of the world from market experience. It does so in the following way. Given its prior belief μ_{i0} firm i sets its (subjective) optimal decision $p_i(\mu_{i0})$. Once each firm has made this move the objective demand density function establishes the quantities that can actually be sold given the actions $p_0 := p_i(\mu_{i0})_{i \in N}$. This means that for each firm i a quantity x_{i1} is drawn from the probability measure that assigns to each Borel set $A \subset X_i$ the probability

$$\int_A f_i(x_i \mid p_0) d\nu_i.$$

This new information is ground for a revision of beliefs via Bayesian updating. Repeating this procedure yields the following learning process.

At a given time $\tau = 0, 1, \ldots$, each individual firm i has recorded a history of consumer demands

$$h_{i\tau} = (x_{it})_{t=1}^{\tau}$$

of finite length τ. This market information is the basis of the belief $\mu_{i\tau}(h_{i\tau})$ of firm i at time τ concerning the state of the world. It then takes a new action $p_i(\mu_{i\tau}(h_{i\tau}))$ based on its current belief. Given the vector $p_\tau := (p_i(\mu_{i\tau}(h_{i\tau})))_{i \in N}$ of new decisions, firm i observes a new quantity $x_{i\tau+1}$ drawn from the probability distribution that assigns to each Borel set $A \subset X_i$ the probability

$$\int_A f_i(x_i \mid p_\tau) d\nu_i.$$

Subsequently beliefs are updated according to the Bayesian updating rule. Formally,

$$\mu_{i\tau+1}(h_{i\tau}, x_{i\tau+1}) := B_i(\mu_{i\tau}(h_{i\tau}))(x_{i\tau+1}).$$

Remarks. Note that the decision on $p_i(\mu_{i\tau}(h_{i\tau}))$ the firm takes at time τ is a function only of the beliefs at time τ, which in turn derive from the initial beliefs μ_{i0} and the recorded history up until τ. Hence, it is sufficient to record sequences of observed quantities, as the firms do.

So we have constructed a well-specified process in which beliefs lead to perceived optimal decisions p_τ, which serve as endogenous signals to obtain new information about the parameters of the distribution of objective demand. This new information, in turn, leads to an update of beliefs and therefore, to new optimal decisions $p_{\tau+1}$. ◁

INFINITE HISTORIES AND BELIEFS

The above-described process driving the decision dynamics of the model thus embodies both subjectively rational learning and subjectively rational actions. In order to study the dynamic properties of this decision process, we make use of martingale convergence theory. For that purpose, we need to construct an underlying probability space on which we can identify martingales. This is the space of all possible future developments of history a firm i foresees at the beginning of time. [3] Formally, let

$$H_{i\tau} := \prod_{t=1}^{\tau} X_i$$

be the space of all histories $h_{i\tau}$ of length τ. $\mathcal{B}(H_{i\tau})$ denotes the Borel σ-algebra on $H_{i\tau}$. Further, let $H_i := \prod_{t=1}^{\infty} X_i$ be the space of infinite histories. A specific element of H_i is denoted by h_i. By $\mathcal{B}(H_i)$ we denote the Borel σ-algebra generated by the product topology on H_i.

To complete the probability space of all future histories, we need a measure λ_i on $\mathcal{B}(H_i)$. Formally this λ_i is defined inductively on histories of finite length, combined with infinite extensions. We will now go through this construction step by step. First note that it is in fact sufficient to specify the numbers

$$\lambda_i(D_\tau \times \prod_{t=\tau+1}^{\infty} X_i)$$

for each Borel set D_τ in $H_{i\tau}$. Because, once these numbers are known, there is a unique way to extend λ_i to $\mathcal{B}(H_i)$. So we only need to specify the numbers

$$\lambda_{i\tau}(\prod_{t=1}^{\tau} D_t),$$

where $\lambda_{i\tau}$ is the probability measure induced by the beliefs of firm i up till time τ. Once these numbers are known, λ_i follows straightforwardly. In fact,

$$\lambda_i(\prod_{t=1}^{\tau} D_t \times \prod_{t=\tau+1}^{\infty} X_i) := \lambda_{i\tau}(\prod_{t=1}^{\tau} D_t),$$

[3] We deviate somewhat from the structure generally chosen for this purpose, *e.g.* in Easley and Kiefer (1988), though in essence the spaces are the same.

the probability that an infinite history starts with a history $h_{i\tau}$ in the set $\prod_{t=1}^{\tau} D_t$. In order to specify these numbers we naturally start with $\lambda_{i0}(\emptyset) := 1$. Further, for $\tau = 1$,

$$\lambda_{i1}(D_1) := \int_{D_1} \int_{\Theta_i} g_i(x_i, \theta_i \mid p_i(\mu_{i0})) d\mu_{i0} d\nu_i.$$

In order to now define $\lambda_{i\tau+1}$ inductively, assume that $\lambda_{i\tau}$ is known. Let $h_{i\tau}$ be a history of length τ. Then the transition probability $\gamma_{i\tau+1}(h_{i\tau})(D_{\tau+1})$ of ending up in $D_{\tau+1} \subset X_i$ provided we have observed history $h_{i\tau}$ is equal to

$$\gamma_{i\tau+1}(h_{i\tau})(D_{\tau+1}) := \int_{D_{\tau+1}} \int_{\Theta_i} g_i(x_i, \theta_i \mid p_i(\mu_{i\tau}(h_{i\tau}))) d\mu_{i\tau}(h_{i\tau}) d\nu_i.$$

The transition probability gives the subjective probability of an observation $x_{i\tau+1}$ being in $D_{\tau+1}$ given that the firm has already observed history $h_{i\tau}$ and subsequently believes that $\mu_{i\tau}(h_{i\tau})$ is the appropriate probability distribution over Θ_i. We then have

$$\begin{aligned}
\lambda_{i\tau+1}(\prod_{t=1}^{\tau+1} D_t) &:= \int_{H_{i\tau}} \int_{X_i} \mathbf{1}_{\prod_{t=1}^{\tau+1} D_t} d\gamma_{i\tau+1}(h_{i\tau}) d\lambda_{i\tau} \\
&= \int_{H_{i\tau}} \mathbf{1}_{\prod_{t=1}^{\tau} D_t} \int_{X_i} \mathbf{1}_{D_{\tau+1}} d\gamma_{i\tau+1}(h_{i\tau}) d\lambda_{i\tau} \\
&= \int_{H_{i\tau}} \mathbf{1}_{\prod_{t=1}^{\tau} D_t} \gamma_{i\tau+1}(h_{i\tau})(D_{\tau+1}) d\lambda_{i\tau} \\
&= \int_{\prod_{t=1}^{\tau} D_t} \gamma_{i\tau+1}(h_{i\tau})(D_{\tau+1}) d\lambda_{i\tau}.
\end{aligned}$$

The definition reflects how $\lambda_{i\tau+1}$ derives as the weighted 'sum' (i.e., the integral) of all transition probabilities, where the weights are the probabilities $\lambda_{i\tau}$ the firm assigns to the observation that conditions the particular transition probability. The first step easily follows from rewriting the indicator function on the product set as a product of indicator functions. It is then observed that the inner integral equals $\gamma_{i\tau+1}(h_{i\tau})(D_{i\tau+1})$. Finally, the indicator function is replaced by the restricted integral.

Now notice that, since the above computation implies that for all sets D_τ in $\mathcal{B}(H_{i\tau})$ we have

$$\lambda_{i\tau+1}(D_\tau \times X_i) = \int_{D_\tau} \gamma_{i\tau+1}(h_{i\tau})(X_i) d\lambda_{i\tau}$$

$$= \int_{D_\tau} \mathbf{1}_{H_{i\tau}} d\lambda_{i\tau} = \lambda_{i\tau}(D_\tau),$$

the measures $\lambda_{i\tau}$ are consistent. Therefore, by the Theorem of Kolmogorov, there is a unique probability measure λ_i on $\mathcal{B}(H_i)$ such that

$$\lambda_i \left(D_\tau \times \prod_{t=\tau+1}^{\infty} X_i \right) = \lambda_{i\tau}(D_\tau).$$

for all Borel sets D_τ in $\mathcal{B}(H_{i\tau})$.

An appealing way to think about λ_i is as the probability firm i initially assigns to observing the infinite history $h_{i\infty} \in H_{i\infty}$, based on its prior beliefs and its awareness of the learning process it is about the engage in. An example may help to clarify this.

Example 1. A stochastic variable X takes on one of two values, x_1 or x_2. The probability of x_1 (and hence x_2) depends on a parameter θ, that is either θ_1 or θ_2. Let $\Pr(x_1, x_2 \mid \theta_1) = (\frac{1}{3}, \frac{2}{3})$ and $\Pr(x_1, x_2 \mid \theta_2) = (\frac{1}{2}, \frac{1}{2})$ be the conditional probabilities of x_1 and x_2, and suppose $\mu_0 = (\frac{1}{4}, \frac{3}{4})$ are the prior beliefs on (θ_1, θ_2). Over time, a sequence of observations $(x_t)_{t \in \mathbb{N}}$ molds beliefs. We have

$$\begin{aligned} \gamma_1(X_1 = x_1) &= \frac{1}{4} \cdot \frac{1}{3} + \frac{3}{4} \cdot \frac{1}{2} = \frac{11}{24} = \lambda_1(X_1 = x_1) \\ \gamma_1(X_1 = x_2) &= \frac{1}{4} \cdot \frac{2}{3} + \frac{3}{4} \cdot \frac{1}{2} = \frac{13}{24} = \lambda_1(X_1 = x_2). \end{aligned}$$

Suppose $X_1 = x_1$. Application of Bayes rule now gives posterior beliefs

$$\mu_1 \mid (X_1 = x_1) = (\frac{\frac{1}{4} \cdot \frac{1}{3}}{\frac{1}{4} \cdot \frac{1}{3} + \frac{3}{4} \cdot \frac{1}{2}}, \frac{\frac{3}{4} \cdot \frac{1}{2}}{\frac{1}{4} \cdot \frac{1}{3} + \frac{3}{4} \cdot \frac{1}{2}}) = (\frac{2}{11}, \frac{9}{11}).$$

Similarly, $X_1 = x_2$ would return

$$\mu_1 \mid (X_1 = x_2) = (\frac{\frac{1}{4} \cdot \frac{2}{3}}{\frac{1}{4} \cdot \frac{2}{3} + \frac{3}{4} \cdot \frac{1}{2}}, \frac{\frac{3}{4} \cdot \frac{1}{2}}{\frac{1}{4} \cdot \frac{2}{3} + \frac{3}{4} \cdot \frac{1}{2}}) = (\frac{4}{13}, \frac{9}{13}).$$

We then have the conditional transition probabilities

$$\begin{aligned} \gamma_2(X_2 = x_1 \mid X_1 = x_1) &= \frac{2}{11} \cdot \frac{1}{3} + \frac{9}{11} \cdot \frac{1}{2} = \frac{31}{66} \\ \gamma_2(X_2 = x_2 \mid X_1 = x_1) &= \frac{2}{11} \cdot \frac{2}{3} + \frac{9}{11} \cdot \frac{1}{2} = \frac{35}{66} \end{aligned}$$

$$\begin{aligned}\gamma_2(X_2 = x_1 \mid X_1 = x_2) &= \frac{4}{13}\cdot\frac{1}{3}+\frac{9}{13}\cdot\frac{1}{2}=\frac{35}{78}\\ \gamma_2(X_2 = x_2 \mid X_1 = x_2) &= \frac{4}{13}\cdot\frac{2}{3}+\frac{9}{13}\cdot\frac{1}{2}=\frac{43}{78}.\end{aligned}$$

The λ-measure for the $t = 2$ paths is now constructed by combining the conditional transition probabilities, as follows.

$$\lambda_2(X_1 = x_1, X_2 = x_1) = \lambda_1(X_1 = x_1)\cdot\gamma_2(X_2 = x_1 \mid X_1 = x_1) = \frac{11}{24}\cdot\frac{31}{66}=\frac{31}{144}.$$

Similarly we find

$$\begin{aligned}\lambda_2(X_1 = x_1, X_2 = x_2) &= \frac{11}{24}\cdot\frac{35}{66}=\frac{35}{144}\\ \lambda_2(X_1 = x_2, X_2 = x_1) &= \frac{13}{24}\cdot\frac{35}{78}=\frac{35}{144}\\ \lambda_2(X_1 = x_2, X_2 = x_2) &= \frac{13}{24}\cdot\frac{43}{78}=\frac{43}{144}.\end{aligned}$$

Finally, the posteriors follow from Bayes' rule as

$$\begin{aligned}\mu_2 \quad &\mid \quad (X_1 = x_1, X_2 = x_1) = (\frac{4}{31},\frac{27}{31})\\ \mu_2 \quad &\mid \quad (X_1 = x_1, X_2 = x_2) = (\frac{8}{35},\frac{27}{35}) = \mu_2 \mid (X_1 = x_2, X_2 = x_1)\\ \mu_2 \quad &\mid \quad (X_1 = x_2, X_2 = x_2) = (\frac{16}{43},\frac{27}{43}).\end{aligned}$$

This concludes the example. ◁

5. Convergence of beliefs and actions

The prime interest in this paper is to know whether, given initial beliefs, the process of Bayesian updating will eventually converge to a conjectural equilibrium. That is, we ask whether learning will teach some invariable posterior ideas, or whether perceptions, and thus decisions, will keep on changing for ever. In order to address this question we will employ a convergence theorem concerning martingales. However, before we can apply this theorem we need to show that, on the probability space $(H_i, \mathcal{B}(H_i), \lambda_i)$ constructed above, beliefs indeed form a martingale. To that end we first need to introduce some notation.

Consider an infinite history $h_i = (x_{it})_{t=1}^{\infty}$ in H_i. The finite history $h_{i\tau} := (x_{it})_{t=1}^{\tau}$ in $H_{i\tau}$ is called the *truncation* of h_i till time τ. Further, let A be a

Borel set in $\mathcal{B}(\Theta_i)$. Consider the function $\mu_{i\tau}(A)$ from H_i to $\mathbb{R}$ that assigns to an infinite history h_i the real number

$$\mu_{i\tau}(A)(h_i) := \mu_{i\tau}(h_{i\tau})(A).$$

Secondly, notice that the above truncation of infinite histories to histories of length τ induces is a natural identification of each element D_τ of the σ-algebra $\mathcal{B}(H_{i\tau})$ with the set

$$D_\tau \times \prod_{t=\tau+1}^{\infty} X_i$$

in $\mathcal{B}(H_i)$. The subalgebra of $\mathcal{B}(H_i)$ of sets of this form is denoted by $\mathcal{B}_\tau(H_i)$.

It is immediately clear that $\mathcal{B}_\tau(H_i)$ is a subset of $\mathcal{B}_{\tau+1}(H_i)$. Furthermore, it is also not so hard to see that each function $\mu_{i\tau}(A)$ is $\mathcal{B}_\tau(H_i)$-measurable and bounded by $K = 1$. In other words, the sequence $(\mu_{i\tau}(A))_{\tau=1}^{\infty}$ provides information [4]. We will show that it is even a martingale.

Theorem 1. *Let A be a Borel set in $\mathcal{B}(\Theta_i)$. Then the sequence $(\mu_{i\tau}(A))_{\tau=1}^{\infty}$ of random variables is a martingale on w.r.t. λ_i.*

Proof. Let A be a Borel set in $\mathcal{B}(\Theta_i)$ and let C be a Borel set in $\mathcal{B}_\tau(H_i)$. We have to check that

$$\int_C \mu_{i\tau+1}(A)(h_i)d\lambda_i = \int_C \mu_{i\tau}(A)(h_i)d\lambda_i.$$

Since C is an element of $\mathcal{B}_\tau(H_i)$ we know it can be written as

$$D_\tau \times \prod_{t=1} X_i$$

for some Borel set D_τ in $H_{i\tau}$. So, since λ_i agrees with $\lambda_{i\tau+1}$ on $\mathcal{B}_{\tau+1}(H_i)$, Lemma 4 in Appendix A yields

$$\int_{D_\tau \times \prod_{t=1} X_i} \mu_{i\tau+1}(A)(h_i)d\lambda_i =$$
$$\int_{D_\tau \times X_i} \mu_{i\tau+1}(h_{i\tau}, x_{i\tau+1})(A)d\lambda_{i\tau+1} =$$
$$\int_{D_\tau} \int_{X_i} \mu_{i\tau+1}(h_{i\tau}, x_{i\tau+1})(A) \int_{\Theta_i} g_i(x_{i\tau+1}, \theta_i \mid p_{i\tau})d\mu_{i\tau}(h_{i\tau})d\nu_i d\lambda_{i\tau}.$$

[4] See Definition 17 in Appendix C.

Plugging Bayes' rule into this expression yields

$$\int_{D_\tau}\int_{X_i}\frac{\int_A g_i(x_{i\tau+1},\theta_i \mid p_{i\tau})d\mu_{i\tau}(h_{i\tau})}{\int_{\Theta_i} g_i(x_{i\tau+1},\theta_i \mid p_{i\tau})d\mu_{i\tau}(h_{i\tau})}\times\int_{\Theta_i} g_i(x_{i\tau+1},\theta_i \mid p_{i\tau})d\mu_{i\tau}(h_{i\tau})d\nu_i d\lambda_{i\tau}$$

and the two integrals over Θ_i cancel out. Which reduces the above expression to

$$\int_{D_\tau}\int_{X_i}\int_A g_i(x_{i\tau+1},\theta_i \mid p_{i\tau})d\mu_{i\tau}(h_{i\tau})d\nu_i d\lambda_{i\tau}.$$

To this expression we can apply the Theorem of Fubini and switch the order of integration over X_i and A. This yields

$$\begin{aligned}\int_{D_\tau}\int_A\int_{X_i} g_i(x_{i\tau+1},\theta_i \mid p_{i\tau})d\nu_i d\mu_{i\tau}(h_{i\tau})d\lambda_{i\tau} &= \int_{D_\tau}\int_A \mathbb{1}_{\Theta_i}d\mu_{i\tau}(h_{i\tau})d\lambda_{i\tau}\\ &= \int_{D_\tau}\mu_{i\tau}(h_{i\tau})(A)d\lambda_{i\tau},\end{aligned}$$

where the first equality results from the fact that g_i is a density function with respect to ν_i. This concludes the proof. ◁

This result may not be very surprising. It states that the nature of Bayesian learning is such that a firm does not expect to change its beliefs in the future. Of course, an actual observation will in general change beliefs, but based upon current beliefs on future realizations of sales, a firm ex ante predicts it will not. One way to interpret this is as Bayesian learning being sufficient, in that the information present at times is used to the full.

Example 2. In our example, it is easy to see that beliefs have the martingale property. The expectation $E_\lambda(\cdot)$ taken with respect to λ is

$$E_\lambda(\mu_1(\theta_1))=\frac{11}{24}\cdot\frac{2}{11}+\frac{13}{24}\cdot\frac{4}{13}=\frac{1}{4}=\mu_0(\theta_1),$$

and similarly

$$E_\lambda(\mu_2(\theta_1))=\frac{31}{144}\cdot\frac{4}{31}+\frac{35}{144}\cdot\frac{8}{35}+\frac{35}{144}\cdot\frac{8}{35}+\frac{43}{144}\cdot\frac{16}{43}=\frac{1}{4}=\mu_0(\theta_1).$$

This concludes the example. ◁

With the result in hand, we can apply the martingale convergence theorem set out in the Appendix. We can use this result to study the limit beliefs of agents, and hence of decisions, as follows.

Take an infinite history h_i in H_i. Let $\mu_{i\tau}(h_i)$ be the probability measure in $\mathbb{P}(\Theta_i)$ that assigns to each Borel set A of Θ_i the real number $\mu_{i\tau}(h_i)(A)$.

Theorem 2. *There exists a Borel set S of infinite histories in H_i with λ_i-probability one on which the sequence $(\mu_{i\tau}(h_i))_{\tau=1}^{\infty}$ of probability measures converges weakly to a probability measure $\mu_{i\infty}(h_i)$ for every history h_i in S.*

Proof. We will first construct S. Since Θ_i is compact and metric, we know that there exists a countable basis of the topology. Let $\mathcal{U}$ be the collection of finite intersections of elements of this basis. Take a fixed element U of $\mathcal{U}$. By Theorem 1, the sequence $(\mu_{i\tau}(U))_{\tau=1}^{\infty}$ is a martingale. So, by Theorem 18 of Appendix C there is a set $S(U)$ of infinite histories in H_i with $\lambda_i(S(U)) = 1$ such that $(\mu_{i\tau}(h_i)(U))_{\tau=1}^{\infty}$ converges for every history h_i in $S(U)$.

Now since $\mathcal{U}$ is the collection of finite intersections of a countable collection, it is a countable set itself. This implies that

$$S := \bigcap_{U \in \mathcal{U}} S(U)$$

has λ_i-probability one, since it is a countable intersection of sets $S(U)$, all having λ_i-probability one.

The construction of the limit probability measure can be done as follows. Take a history h_i in S. Since $\mathbb{P}(\Theta_i)$ is sequentially compact by Theorems 12, 13 and 14 of Appendix B, we know that a subsequence of $(\mu_{i\tau}(h_i))_{\tau=1}^{\infty}$ converges weakly to some probability measure, say $\mu_{i\infty}(h_i)$. We will show that the original sequence converges weakly to this probability measure. To this end, notice that

$$\mu_{i\tau}(h_i)(U) \to \mu_{i\infty}(h_i)(U) \quad \text{for all } U \in \mathcal{U}$$

for the original sequence, since this sequence is convergent for every element U of $\mathcal{U}$ by construction of S and the above holds for the weakly convergent subsequence. Moreover, $\mathcal{U}$ is closed under finite intersections and each open set is obviously a countable union of elements of $\mathcal{U}$ since $\mathcal{U}$ contains a countable basis of the topology on Θ_i by construction. Hence, by Lemma 8, $(\mu_{i\tau}(h_i))_{\tau=1}^{\infty}$

converges weakly to $\mu_{i\infty}(h_i)$ and the proof is complete since h_i was chosen arbitrarily in S. ◁

From now on we will automatically assume that we only consider histories h_i in S whenever we talk about $\mu_{i\infty}(h_i)$. Effectively, we only consider the domain of $\mu_{i\infty}$. We can now prove the following result.

Theorem 3. *The sequence $p_i(\mu_{i\tau}(h_i))_{\tau=1}^{\infty}$ of actions λ_i-almost-surely converges to the limit decision $p_i(\mu_{i\infty}(h_i))$.*

Proof. By the continuity of p_i established in Lemma 14 of Appendix D, we know that the sequence $p_i(\mu_{i\tau}(h_i))_{\tau=1}^{\infty}$ of optimal decisions given beliefs at time τ converges to $p_i(\mu_{i\infty}(h_i))$ whenever the sequence $\mu_{i\tau}(h_i)_{\tau=1}^{\infty}$ of beliefs converges to $\mu_{i\infty}(h_i)$. This though happens with λ_i-probability one by Theorem 2. ◁

6. The nature of limit beliefs and limit actions

We now know that in our model beliefs, and consequently decisions, converge to limit beliefs and unique limit decisions respectively, for λ_i-almost-all developments of history. In this section we will derive some properties of the limit beliefs and decisions. We will show that a limit belief is unique in the sense that, roughly speaking, it only puts weight on worlds that generate the same probability distribution over demands. Furthermore we will show that it supports a conjectural equilibrium.

UNIQUE LIMIT BELIEFS

For an analysis of the limit properties of beliefs and decisions, consider the following construction. Let μ_i be a probability measure on Θ_i. Evidently Θ_i is a compact set with $\mu_i(\Theta_i) = 1$. So, the collection

$$\mathcal{K} := \{K \subset \Theta_i \mid K \text{ is compact and } \mu_i(K) = 1\}$$

is not empty. Thus we can define the support of μ_i by

$$\mathrm{supp}(\mu_i) := \bigcap_{K \in \mathcal{K}} K.$$

The only question is whether this set has probability one according to μ_i. To this end, notice that the topology on Θ_i has a countable basis, say $\mathcal{B}$, since Θ_i is separable and metric. So,

$$\text{supp}(\mu_i) = \bigcap_{B \in \mathcal{B}: \mu_i(B)=0} \Theta_i \setminus B.$$

Hence, $\mu_i(\text{supp}(\mu_i)) = 1$ by the subadditivity of μ_i.

A more colloquial definition of the support of a probability measure μ_i on Θ_i is to say that it is the smallest compact subset K of Θ_i with $\mu_i(K) = 1$. Anyhow, it enables us to give the following

Definition 3. A belief μ_i *does not distinguish* if there exists a function $h_i : X_i \to \mathbb{R}$, such that for any θ_i in $\text{supp}(\mu_i)$ and for all x_i in X_i

$$g_i(x_i \mid p_i(\mu_i), \theta_i) = h_i(x_i).$$

This condition on μ_i states that every world θ_i in the support of μ_i generates the same density function on X_i. Consequently, no signal x_i will give firm i a reason to change its belief. A more interesting fact is that the converse of this observation is also true. This is reflected in

Theorem 4. *A belief μ_i does not distinguish if and only if*

$$B_i(\mu_i)(x_i) = \mu_i$$

holds for all x_i in X_i.

Proof. Suppose that μ_i does not distinguish. Then we can take $h_i : X_i \to \mathbb{R}$, such that

$$h_i(x_i) = g_i(x_i \mid p_i(\mu_i), \theta_i) \quad \text{for all } \theta_i \in \text{supp}(\mu_i).$$

Consequently, for any $x_i \in X_i$, and any Borel set A in Θ_i we have

$$\begin{aligned} B_i(\mu_i)(x_i)(A) = \frac{\int_A g_i(x_i \mid p_i(\mu_i), \theta_i) d\mu_i}{\int_{\Theta_i} g_i(x_i \mid p_i(\mu_i), \theta_i) d\mu_i} &= \frac{\int_A h_i(x_i) \mathbf{1}_{\text{supp}(\mu_i)} d\mu_i}{\int_{\Theta_i} h_i(x_i) \mathbf{1}_{\text{supp}(\mu_i)} d\mu_i} \\ &= \frac{h_i(x_i)\mu_i(A)}{h_i(x_i)\mu_i(\Theta_i)} = \mu_i(A). \end{aligned}$$

Suppose, on the other hand, that μ_i distinguishes. Then we know that there is a pair $\zeta_i, \gamma_i \in \text{supp}(\mu_i)$, and an $x_i^* \in X_i$ for which

$$g_i(x_i^* \mid p_i(\mu_i), \zeta_i) > g_i(x_i^* \mid p_i(\mu_i), \gamma_i).$$

So we can find two positive numbers $U > L \in \mathbb{R}$ and open neighborhoods $N(\zeta_i) \ni \zeta_i$ and $N(\gamma_i) \ni \gamma_i$ such that for all θ_i in $N(\zeta_i)$

$$g_i(x_i^* \mid p_i(\mu_i), \theta_i) \geq U$$

and for all θ_i in $N(\gamma_i)$

$$g_i(x_i^* \mid p_i(\mu_i), \theta_i) \leq L.$$

Now notice that $\mu_i(N(\zeta_i)) > 0$ since otherwise $\text{supp}(\mu_i) \setminus N(\zeta_i)$ would be a compact set with μ_i-probability one that is strictly included in $\text{supp}(\mu_i)$. For the same reason $\mu_i(N(\gamma_i)) > 0$. This implies that

$$\frac{B_i(\mu_i)(x_i^*)(N(\zeta_i))}{B_i(\mu_i)(x_i^*)(N(\gamma_i))} \geq \frac{\int_{N(\zeta_i)} U \mathbf{1}_{\Theta_i} d\mu_i}{\int_{N(\gamma_i)} L \mathbf{1}_{\Theta_i} d\mu_i} = \frac{U \mu_i(N(\zeta_i))}{L \mu_i(N(\gamma_i))} > \frac{\mu_i(N(\zeta_i))}{\mu_i(N(\gamma_i))}.$$

So, at least

$$B_i(\mu_i)(x_i^*)(N(\zeta_i)) \neq \mu_i(N(\zeta_i))$$

or

$$B_i(\mu_i)(x_i^*)(N(\gamma_i)) \neq \mu_i(N(\gamma_i)).$$

In any case, $B_i(\mu_i)(x_i^*)$ does not equal μ_i and the proof is complete. ◁

The interpretation of the proposition is straightforward. A belief μ_i does not distinguish if and only if Bayesian updating has no effect on the belief for any possible signal x_i. This fact has important implications. Particularly since we can show that the limit beliefs $\mu_{i\infty}(h_i)$ in fact are fixed points of the Bayesian updating method as we will do now.

To this end, we need the following preliminary result. Let $\mathcal{B}$ be a countable basis of the topology on X_i. Let W be the collection of sample paths $(x_{it})_{t=1}^{\infty}$ in H_i for which there is a basis element B in $\mathcal{B}$ such that $\{x_{it} \mid x_{it} \in B\}$ is finite. We will show first that the following is true.

Lemma 1. $\lambda_i(W) = 0$.

Proof. Let B be an element of $\mathcal{B}$ and let T be a natural number. Define

$$W(B,T) := \{(x_{it})_{t=1}^{\infty} \mid x_{it} \notin B \text{ for all } t \geq T\}.$$

Note that this construction is such that $W = \bigcup_{B,T} W(B,T)$. So, W is the countable union of sets $W(B,T)$. Hence, by the subadditivity of λ_i it suffices to prove that $\lambda_i(W(B,T)) = 0$ for any choice of B and T.

To this end, notice that

$$W(B,T) = \prod_{t=1}^{T} X_i \times \prod_{t=T+1}^{\infty} B^c.$$

Now take some $\tau \geq T$. Denote the subset $\prod_{t=1}^{T} X_i \times \prod_{t=T+1}^{\tau} B^c$ of the set $H_{i\tau}$ of finite histories up till time τ by W_τ. Then, for a history $h_{i\tau}$ in W_τ, the one-step transition probability $\gamma_{i\tau+1}(h_{i\tau}(B)$ to B is

$$\begin{aligned}\gamma_{i\tau+1}(h_{i\tau})(B) \quad &:= \quad \int_B \int_{\Theta_i} g_i(x_i, \theta_i \mid p_{i\tau}) d\mu_{i\tau}(h_{i\tau}) d\nu_i \\ &\geq \quad \int_B \int_{\Theta_i} \varepsilon d\mu_{i\tau}(h_{i\tau}) d\nu_i = \varepsilon \nu_i(B).\end{aligned}$$

Here $\varepsilon > 0$ is chosen such that $g_i(x_i, \theta_i \mid p_{i\tau}) \geq \varepsilon$ for all x_i and θ_i, which can be done by the compactness of X_i, the continuity of g_i and the assumption that g_i is larger than zero on X_i. Consequently, $\gamma_{i\tau+1}(h_{i\tau})(B^c) \leq 1 - \varepsilon\nu_i(B)$. Using this result, we get that

$$\lambda_{i\tau+1}(W_{\tau+1}) := \int_{W_\tau} \gamma_{i\tau+1}(h_{i\tau})(B^c) d\lambda_{i\tau} \leq (1 - \varepsilon\nu_i(B))\lambda_{i\tau}(W_\tau).$$

Now backsubstitution yields

$$\lambda_{i\tau+1}(W_{\tau+1}) \leq (1 - \varepsilon\nu_i(B))^{\tau - T + 1} \lambda_{iT}(W_T) = (1 - \varepsilon\nu_i(B))^{\tau - T + 1}.$$

Further, since B is an open set, we know that $\nu_i(B) > 0$ by assumption. So, $0 \leq 1 - \varepsilon\nu_i(B) < 1$ and hence

$$\lim_{\tau \to \infty} \lambda_{i\tau+1}(W_{\tau+1}) = 0.$$

Finally, since $0 \leq \lambda_i(W) \leq \lambda_{i\tau}(W_\tau)$ for all $\tau \geq T$ by construction of λ_i, it follows that $\lambda_i(W) = 0$. ◁

The interpretation of this result is that firms expect *a priori* that the signals they will receive are persistently exciting. That is, they expect to observe all possible quantities infinitely many times over the course of their learning process, so that they will be able to indeed extract sufficient information from them. The sufficiency of the information is reflected in

Theorem 5. *There is a subset Z of S with λ_i-probability one such that the belief $\mu_{i\infty}(h_i)$ does not distinguish for any h_i in Z.*

Proof. Let S be as in Theorem 2 and let W be as in Lemma 1. Write $Z := S \setminus W$. Clearly, $\lambda_i(Z) = 1$, since $\lambda_i(S) = 1$ and $\lambda_i(W) = 0$. Now take a history $h_i = (x_{i\tau})_{\tau=1}^{\infty}$ in Z. Then, since h_i is an element of S we know that $\mu_{i\infty}(h_i)$ exists. We will show that it does not distinguish.

By Theorem 4 it suffices to show that $B(\mu_{i\infty}(h_i))(x_i) = \mu_{i\infty}(h_i)$ for all x_i in X_i. To this end, take an $x_i^* \in X_i$. Then, since $h_i = (x_{i\tau})_{\tau=1}^{\infty}$ is not an element of W, we know that it intersects each element of the basis $\mathcal{B}$ infinitely many times. So, since X_i is metric, this implies that we can find a subsequence $(x_{i\alpha(\tau)})_{\tau=1}^{\infty}$ of $(x_{i\tau})_{\tau=1}^{\infty}$ such that $x_{i\alpha(\tau)} \to x_i^*$ as $\tau \to \infty$. Then, on one hand

$$B(\mu_{i\alpha(\tau)}(h_{i\alpha(\tau)}))(x_{i\alpha(\tau)+1}) = \mu_{i\alpha(\tau)+1}(h_{i\alpha(\tau)+1}) = \mu_{i\alpha(\tau)+1}(h_i) \to \mu_{i\infty}(h_i)$$

in the weak topology since the above sequence is a subsequence of $(\mu_{i\tau}(h_i))_{\tau=1}^{\infty}$ which converges to $\mu_{i\infty}(h_i)$ in the weak topology by the choice of S. On the other hand,

$$B(\mu_{i\alpha(\tau)}(h_{i\alpha(\tau)}))(x_{i\alpha(\tau)+1}) \to B(\mu_{i\infty}(h_i))(x_i^*)$$

since B is continuous by Theorem 19 of Appendix D. Hence, since the space $\mathbb{P}(\Theta_i)$ is Hausdorff, $\mu_{i\infty}(h_i) = B(\mu_{i\infty}(h_i))(x^*)$. ◁

Note that if we make the natural assumption that conjectured density functions of demand are uniquely characterized by the value of θ_i, the proposition implies that the posterior distribution converges to a point mass on one particular world θ_i in Θ_i.

Assumption 3. For any $p_i \in P_i$ we have $g_i(x_i \mid p_i, \zeta_i) = g_i(x_i \mid p_i, \gamma_i)$ for all $x_i \in X_i$ if and only if $\zeta_i = \gamma_i$.

For a world θ_i the measure that puts probability one on θ_i is called a Dirac measure or a point mass. We have the following result.

Corollary 1. *Suppose we have Assumption 3. Then $\mu_{i\infty}(h_i)$ is a Dirac measure for every h_i in Z.*

Proof. Let h_i be a history in Z. Then $\mu_{i\infty}(h_i)$ does not distinguish by Theorem 5. So, for any pair of worlds ζ_i and γ_i in the support of $\mu_{i\infty}(h_i)$ we have that

$$g_i(x_i \mid p_i(h_i), \zeta_i) = g_i(x_i \mid p_i(h_i), \gamma_i)$$

for the unique limit decision $p_i(h_i) := p_i(\mu_{i\infty}(h_i))$ in P_i and all x_i in X_i. Further, by Assumption 3, this can only be the case if $\zeta_i = \gamma_i$. Hence, the support of $\mu_{i\infty}(h_i)$ is inevitably a singleton and $\mu_{i\infty}(h_i)$ is a Dirac measure. ◃

CONJECTURAL EQUILIBRIUM

Provided the structure of perceptions satisfies Assumptions 1-3 we have shown that, with λ_i-probability one, firm i's belief is a Dirac measure $\mu_{i\infty}(h_i)$. Consequently, firm i's limit decision is $p_i(h_i) := p_i(\mu_{i\infty}(h_i))$. Let $\theta_i(h_i)$ be the unique world in the support of $\mu_{i\infty}(h_i)$. The pair $(\theta_i(h_i), p_i(h_i))$ then specifies the limit stochastic view of the world of each firm. That is, each firm i perceives demand to be distributed in the limit as

$$g_i(x_i \mid p_i(h_i), \theta_i(h_i)).$$

We can now relate our results straightforwardly with our concept of equilibrium. We say that convergence is *almost sure* if it is λ_i-almost sure for every i.

Theorem 6. *The learning process almost surely converges to a conjectural equilibrium.*

Proof. By Theorem 5 we know that the belief $\mu_{i\infty}(h_i)$ of firm i does not distinguish on Z. So, by Theorem 4 it is a fixed point of the Bayes operator and hence an individual conjectural equilibrium. Since this holds for every firm these beliefs form a conjectural equilibrium. ◃

7. Objective Convergence to Conjectural Equilibrium

We now know that for almost all developments of history to which a firm initially assigns non-zero probability, its beliefs on the parameters of conjectured market demand, and thereby the decisions it takes, converge to a unique limit belief that puts all mass on a single parameter of conjectured demand. For each firm, the limit decision is an individual conjectural equilibrium.

Since these results hold for every individual firm i, we are indeed close to conclusions on the behaviour of the complete economy. However, since the conjectures that firms entertain are structurally misspecified, their beliefs of possible developments of history need not necessarily match with the objective sequence of market demand they face. Consequently, actual histories may unfold that have λ_i-probability zero for some firms. Firms facing such probability zero histories will be unable to cope with it: Bayesian learning breaks down under such shocking surprises, and convergence fails. In order to exclude the rise of such paths, therefore, we need a condition that relates beliefs to objective probabilities.

The objective probability measure on the space of sample paths of the form $h_i \in H_i$ is potentially influenced by the behaviour of all firms through the objective demand functions $f_i(x_i \mid p)$. In fact, for given initial beliefs μ_0 of the population, the unfolding sequence of individual actions that derives from the firms' sequential individual application of Bayes' rule within their conjectured demand structures, lays out a complete history of the world, when performed in the interrelated objective demand structures. For given priors, the only stochastic influence on the individually observed history h_i is from $f_i(x_i \mid p_\tau)$ for each τ.

The construction of objective probabilities on space H_i requires an objective probability measure ρ_i on $\mathcal{B}(H_i)$. Like λ_i, ρ_i is formally defined inductively on histories of finite length, combined with infinite extensions. For $\tau = 0$ we naturally have $\rho_{i0}(\emptyset) = 1$. In order to now define $\rho_{i,\tau+1}$ inductively, assume that $\rho_{i\tau}$ is known. Then, given that we have a history $h_{i\tau}$ of length τ, we can define the transition probability $\delta_{i\tau+1}(h_{i\tau})(D_{\tau+1})$ for each Borel subset $D_{\tau+1}$ of X_i as

$$\delta_{i\tau+1}(h_{i\tau})(D_{\tau+1}) = \int_{D_{\tau+1}} f_i(x_i|p_\tau)d\nu_i.$$

Again we can define

$$\rho_{i\tau+1}(\prod_{t=1}^{\tau+1} D_t) := \int_{\times_{t=1}^{\tau} D_t} \delta_{i\tau+1}(h_{i\tau})(D_{\tau+1})d\rho_{i\tau}$$

and apply the Theorem of Kolmogorov. We now come to a crucial relationship between the objective and subjective probability measures ρ_i and λ_i.

Assumption 4. The probability measure ρ_i is absolutely continuous with respect to probability measure λ_i for every firm i.

In the interpretation that we have offered for the measures λ_i and ρ_i, absolute continuity of ρ_i with respect to λ_i implies that no actual development is possible that was not *a priori* foreseen as a possibility by the firm concerned. There is, therefore, 'No Statistical Surprise' on the side of firms. This may seem strong, but is an assumption that it in fact often (implicitly) made in econometric specifications. Moreover, it seems a natural condition necessary for beliefs to settle down, as one can hardly expect beliefs to converge if all the time new and unforeseen events stir up the learning process. We make it, and then have the following prime result.

Theorem 7. *Beliefs almost surely converge to a conjectural equilibrium.*

Proof. By Theorem 6 we have that the beliefs of each firm i converge to an individual conjectural equilibrium λ_i-almost surely. Since ρ_i is absolutely continuous with respect to λ_i, this convergence is also ρ_i-almost-surely. $\triangleleft$

Again, since p_i is continuous, we get

Theorem 8. *The decision vector* p_τ *converges* ρ_i*-almost surely to a conjectural equilibrium decision vector* p_∞.

References

Arrow, K.J., "Toward a Theory of Price Adjustment," in: Abramovitz, M. et. al. (eds.), *The Allocation of Economic Resources*, Stanford University Press, Stanford, 1959.

Billingsley, P., "Convergence of probability measures", John Wiley and Sons, New York, 1968;

Blume, L., Easley, D., "Learning to be Rational," *Journal of Economic Theory*, 26, 1982.

—, "Rational Expectations Equilibrium: An Alternative Approach," *Journal of Economic Theory*, 34, 1986.

Bray, M., Savin, N.E., "Rational Expectations Equilibria, Learning, and Model Specification," *Econometrica*, 54, 1986.

Bray, M., Kreps, D., "Rational Learning and Rational Expectations," in: G. Feiwell (ed.), *Arrow and the Ascent of Modern Economic Theory*, New York University Press, New York, 1987.

Easley, D., Kiefer, N.M., "Controlling a Stochasitic Process with Unknown Parameters," Econometrica, 56, 1988.

Fisher, F.M., *Disequilibrium Foundations of Equilibrium Economics*, Cambridge University Press, Cambridge, 1983.

Hahn, F., "Information Dynamics and Equilibrium," in: F. Hahn (ed.), *The Economics of Missing Markets, Information, and Games*, Clarendon Press, Oxford, 1989.

Kalai, E., Lehrer, E., "Rational Learning Leads to Nash Equilibrium," *Econometrica*, 61 (5), 1993.

—, "Subjective Games and Equilibria," *Games and Economic Behaviour*, 8, 1995.

Kirman, A.P., "Learning by Firms about Demand Conditions," in: R. Day (ed.), *Adaptive Economics*, Academic Press, New York, 1975.

—, "Mistaken Beliefs and Resultant Equilibria," in: R. Frydman and E. Phelps (eds.), *Individual Forecasting and Collective Outcomes*, Cambridge University Press, Cambridge, 1983.

—, "Learning in Oligopoly: Theory, Simulation, and Experimental Evidence," in: P.A. Kirman and M. Salmon (eds.), *Learning and Rationality in Economics*, Blackwell, Oxford, 1995.

Kolmogorov, A.N., Fomin, S.V., *Introductory Real Analysis*, (transl. R.A. Silverman), Dover Publications, New York, 1970.

Nyarko, Y., "Learning in Mis-Specified Models and the Possibility of Cycles," *Journal of Economic Theory*, 55, 1991.

In the Appendices we collected most of the theoretical framework needed in the paper. We did not try to make it completely self-contained. Nevertheless, anyone with a basic knowledge of topology and measure theory should be able to understand all of it. Mainly we tried to build the theory along the shortest route possible. Most of the theory presented here can be found in some form in a number of textbooks such as Billingsley (1968) or Kolomogorov and Fomin (1970). Usually however you also need to have read at least half of these books before you are able to understand the proofs of the theorems we need. Therefore we decided to include these Appendices in order to give the reader the opportunity to require the insights needed in the paper as quickly as possible.

Appendix A provides some basic definitions concerning probability measures as well as a short treaty on regularity of probability measures and some immediate consequences thereof. Appendix B treats the notion of weak convergence and various alternative descriptions of its related topology. Appendix C is basically a complete proof of (a simple version of) the martingale convergence theorem, taking only the Radon-Nikodym Theorem as given. The proofs in these first three Appendices are mainly based on Billingsley (1968). Appendix D is completely geared towards the paper itself and provides a detailed and complete proof of the continuity of the Bayes operator. Finally, Appendix E provides proofs concerning the support of some of the probability measures used in the paper. The latter two Appendices are based on Easley and Kiefer (1988).

APPENDIX A. PROBABILITY MEASURES

In this Appendix we provide some basic measure theoretic notions as well as a treatment of regularity probability measures and some of its consequnces. Before we can introduce the concept of (probability) measures, we need the notion of an algebra. Suppose we have a (non-empty) set X.

Definition 4. (algebra) A collection $\aleph$ of subsets of X is called an *algebra* if:
(i) $\emptyset \in \aleph$
(ii) if $A \in \aleph$ then $X \setminus A \in \aleph$ and
(iii) if $A_1, \ldots, A_n$ are elements of $\aleph$, then $\cup_{i=1}^{n} A_i \in \aleph$.

Conditions (ii) and (iii) automatically imply that finite intersections of elements of $\aleph$ are also elements of $\aleph$.

Definition 5. (σ-algebra) An algebra $\aleph$ is called a σ*-algebra* if it moreover holds that:
(iv) if $A_1, A_2, \ldots$ is a countable sequence of elements of $\aleph$, then $\cup_{i=1}^{\infty} A_i$ is also an element of $\aleph$.

A sequence $A_1, A_2, \ldots$ is called *mutually disjoint* (m.d. for short) if the intersection of A_i and A_j is empty whenever i is not equal to j. Now let $\aleph$ be a σ-algebra on X. The central notion of measure theory is

Definition 6. (measure) A non-negative function

$$\mu : \aleph \to \mathbb{R}$$

is called a *measure* if for every m.d. sequence $A_1, A_2, \ldots$ in $\aleph$ it holds that

$$\mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i).$$

It goes without saying that the expression on the right hand side of the equality is supposed to exist. The condition itself is called the *σ-additivity* of μ.

Definition 7. (probability measure) A measure μ with $\mu(X) = 1$ is called a *probability measure.*

Suppose that we have a topology τ on X. With this topology we can associate a σ-algebra on X in a very natural way. To see this, first notice that the collection 2^X of all subsets of X is a σ-algebra that contains τ. So, the collection $\mathcal{V}$ of all σ-algebra's that contain τ is not empty. This means that

$$\mathcal{B} := \cap_{\aleph \in \mathcal{V}} \aleph$$

is a non-empty collection of subsets of X. Even better, it is a σ-algebra that, evident by construction, contains τ. We say that τ *generates* this σ-algebra.

Definition 8. (Borel σ-algebra) Any σ-algebra that is generated by a topology is called a *Borel* σ-algebra.

REGULARITY

Let $\mathcal{B}$ be the Borel σ-algebra associated with a metric space (X, d) and let μ be a probability measure on $\mathcal{B}$.

Definition 9. (regularity) We say that μ is *regular* if for every Borel set A in $\mathcal{B}$ and every real number $\varepsilon > 0$ we can find a closed set F and an open set U such that $F \subset A \subset U$ and $\mu(U \setminus F) < \varepsilon$.

That is, a measure is regular if every Borel set can be enclosed by an open set, and can itself enclose a closed set, such that the measure of the difference between the sandwiching sets is arbitrarily close to zero.

Theorem 9. *Every probability measure μ on $\mathcal{B}$ is regular.*

Proof. Let μ be an arbitrary probability measure on Θ. Let $\mathcal{R}$ defined as the collection of sets $A \subset X$ for which for every $\varepsilon > 0$ there exist a closed set F and an open set U such that

$$F \subset A \subset U \text{ and } \mu(U \setminus F) < \varepsilon.$$

Notice that μ is regular if and only if $\mathcal{B}$ is a subset of $\mathcal{R}$. Now, in order to show that $\mathcal{B}$ is indeed a subset of $\mathcal{R}$ we make two steps. First of all we will show that any closed set is an element of $\mathcal{R}$. Then we will show that $\mathcal{R}$ is a σ-algebra. Since $\mathcal{B}$ is by definition the smallest σ-algebra that contains all open, and thus also all closed, sets these two facts together imply that $\mathcal{B}$ is a subset of $\mathcal{R}$ and the proof is complete.

Step 1. Take an arbitrary closed set A. We will show that it is an element of $\mathcal{R}$. To this end, take a real number $\varepsilon > 0$. We will construct F and U. Since A is closed we can simply take $F := A$. In order to construct U, define for each natural number n the open set

$$U_n := \left\{ x \in X \mid d(x, A) < \frac{1}{n} \right\}$$

where $d(x, A) := \inf \{d(x, a) \mid a \in A\}$. It is readily seen that $U_1 \supset U_2 \supset \ldots$. Moreover, $A = \cap_{n=1}^{\infty} U_n$ since A is closed. Now define $R_1 := X \setminus U_1$ and

$$R_n := U_{n-1} \setminus U_n$$

for $n \geq 2$. Then $R_1, R_2, \ldots$ are mutually disjoint since $U_1 \supset U_2 \supset \ldots$. Moreover,

$$\bigcup_{n=1}^{\infty} R_n = (X \setminus U_1) \cup \bigcup_{n=2}^{\infty} U_{n-1} \setminus U_n = X \setminus \cap_{n=1}^{\infty} U_n = X \setminus A.$$

So, since all sets involved are clearly Borel sets and μ is σ-additive,

$$1 - \mu(A) = \mu(X \setminus A) = \mu(\bigcup_{n=1}^{\infty} R_n) = \sum_{n=1}^{\infty} \mu(R_n).$$

Hence, given the chosen $\varepsilon > 0$, we can take a natural number N_ε such that

$$\left| \sum_{n=1}^{N_\varepsilon} \mu(R_n) - (1 - \mu(A)) \right| < \varepsilon.$$

Now define $U := U_{N\varepsilon}$. Then on one hand U is an open set that, by the definition of $U = U_{N\varepsilon}$, clearly contains A. On the other hand we get that

$$\bigcup_{n=1}^{N_\varepsilon} R_n = (X \setminus U_1) \cup (U_1 \setminus U_2) \cup \ldots \cup (U_{N_\varepsilon - 1} \setminus U_{N_\varepsilon}) = X \setminus U_{N_\varepsilon}.$$

Hence,

$$\begin{aligned} \mu(U \setminus F) &= |\mu(U \setminus F)| = |\mu(U) - \mu(F)| = |\mu(U_{N_\varepsilon}) - \mu(A)| \\ &= |(1 - \mu(U_{N_\varepsilon}) - (1 - \mu(A))| = |\mu(X \setminus U_{N_\varepsilon}) - (1 - \mu(A))| \\ &= \left|\mu(\cup_{n=1}^{N_\varepsilon} R_n) - (1 - \mu(A))\right| = \left|\sum_{n=1}^{N_\varepsilon} \mu(R_n) - (1 - \mu(A))\right| < \varepsilon, \end{aligned}$$

where the second equality holds since $F \subset U$. This shows that A is indeed an element of $\mathcal{R}$.

Step 2. Now we will show that $\mathcal{R}$ is a σ-algebra. The first requirement of the definition of a σ-algebra is easy to check since it follows from the previous step that the empty set is an element of $\mathcal{R}$.

Part A. Next we have to show that $X \setminus A$ is an element of $\mathcal{R}$ for every element A of $\mathcal{R}$. We will even show a somewhat stronger statement, namely that $A \setminus B$ is an element of $\mathcal{R}$ for any two sets A and B in $\mathcal{R}$.

So, take two sets A and B in $\mathcal{R}$. Then we can take open sets U and V and closed sets F and G with

$$F \subset A \subset U \quad \text{and} \quad G \subset B \subset V$$

such that $$\mu(U \setminus F) < \varepsilon \quad \text{and} \quad \mu(V \setminus G) < \varepsilon.$$

Define $H := F \setminus V$ and $W := U \setminus G$. It is easy to check that W is open, H is closed and $H \subset A \setminus B \subset W$. Finally, it is elementary to show that $W \setminus H$ is a subset of the union of $U \setminus F$ and $V \setminus G$. Hence,

$$\mu(W \setminus H) \leq \mu(U \setminus F) + \mu(V \setminus G) < \varepsilon + \varepsilon = 2\varepsilon.$$

So, at least we know now that $X \setminus A$ is an element of $\mathcal{R}$ for every A in $\mathcal{R}$.

Part B. To get the third requirement, let $A_1, A_2, \ldots$ be a sequence in $\mathcal{R}$. We have to show that $A := \cup_n A_n$ is also an element of $\mathcal{R}$. This we will also do in two steps. In this first step we make the additional assumption that the sequence is mutually disjoint. Now take a real number $\varepsilon > 0$. Since A_n is an element of $\mathcal{R}$, we can take an open set U_n and a closed set F_n such that

$$F_n \subset A_n \subset U_n \quad \text{and} \quad \mu(U_n \setminus F_n) < \left(\frac{1}{2}\right)^n \varepsilon.$$

Since the sequence $A_1, A_2, \ldots$ is mutually disjoint, it is clear that the sequence $F_1, F_2, \ldots$ is also mutually disjoint. So, by the σ-additivity of μ we know that

$$\sum_{n=1}^{\infty} \mu(F_n)$$

exists and we can take an N such that

$$\sum_{n=N+1}^{\infty} \mu(F_n) < \varepsilon.$$

Take $U := \cup_n U_n$ and $F := \cup_{n=1}^{N} F_n$. Clearly, U is open and F is closed, while $F \subset A \subset U$. Moreover, $U \setminus F$ is a subset of the union of $U_1 \setminus F_1, U_2 \setminus F_2, \ldots$ together with $F_{N+1}, F_{N+2}, \ldots$. This however implies that $\mu(U \setminus F)$ is less than or equal to

$$\sum_{n=1}^{\infty} \mu(U_n \setminus F_n) + \sum_{n=N+1}^{\infty} \mu(F_n) = \sum_{n=1}^{\infty} (\frac{1}{2})^n \varepsilon + \varepsilon = 2\varepsilon.$$

Part C. Now in the third part we will show that the union A of a sequence $A_1, A_2, \ldots$ of elements of $\mathcal{R}$ can be written as the union of a mutually disjoint sequence $B_1, B_2, \ldots$ of elements of $\mathcal{R}$. Then from the above argument in Part B we can conclude that A is indeed an element of $\mathcal{R}$ and the proof is complete.

To this end, define the sequence $B_1, B_2, \ldots$ as follows. Take $B_1 := A_1$ and define B_n recursively by

$$B_n := A_n \setminus B_{n-1}.$$

Obviously the sequence is mutually disjoint. Furthermore, $B_1 = A_1$ is clearly an element of $\mathcal{R}$. So, since A_2 is also an element of $\mathcal{R}$ we know by our result in Part A that $B_2 = A_2 \setminus B_1$ is also an element of $\mathcal{R}$. Hence, by iterating this argument we get that every B_n is an element of $\mathcal{R}$. ◃

The fact that a measure μ on a metric space X is regular has some nice consequences, especially when X is compact. We will discuss some of them.

For a set A in X and a real number $\varepsilon > 0$, write $A_\varepsilon := \{x \in X \mid d(x, A) < \varepsilon\}$ and $A^\varepsilon := \{x \in X \mid d(x, A) \leq \varepsilon\}$. Further, when $A = \{x\}$ we will write x_ε and x^ε instead of $\{x\}_\varepsilon$ and $\{x\}^\varepsilon$. The boundary ∂A of A is the set

$$\partial A := \{x \in X \mid \text{ for every } \varepsilon > 0, x_\varepsilon \cap A \neq \emptyset \text{ and } x_\varepsilon \cap A^c \neq \emptyset\}.$$

The interior $\text{int}(A)$ is defined as $A \setminus \partial A$. Note that ∂A is closed and $\text{int}(A)$ is open. Both sets are therefore elements of $\mathcal{B}$, no matter what A is. Now let μ be a probability measure on X. Then A is called *μ-continuous* if $\mu(\partial A) = 0$, that is if the boundary of A has μ-probability zero. We have the following three results.

Lemma 2. *Let F be a closed set in X and let $\eta > 0$. There exists an $\varepsilon > 0$ such that*

$$\mu(F^\varepsilon) - \mu(F) < \eta.$$

The same inequality automatically holds for all $\delta < \varepsilon$ and F_ε instead of F^ε.

Proof. Take an $\eta > 0$. Since F is closed, regularity of μ implies that there is an open set $U \supset F$ such that $\mu(U \setminus F) < \eta$. We will show that there is a natural number n such that $F^{\frac{1}{n}}$ is a subset of U.

So, suppose that this is not the case. Then we can find a point x_n in $F^{\frac{1}{n}} \setminus U$ for every n. Since X is compact, we may assume w.l.o.g. that this sequence of points has a limit, say x. Then, since $d(x, F) = 0$ and F is closed, x must be an element of F. On the other, all x_n lie outside U and U is open. So, x is not an element of U. This contradicts the assumption that F is included in U. ◃

Corollary 2. *Let A be a μ-continuous Borel set in X and let $\eta > 0$. Then for all sufficiently small $\varepsilon > 0$,*

$$\mu(A^\varepsilon) - \mu(A) < \eta.$$

Proof. The corollary follows easily from the previous Lemma once we have made the observations that, since A is μ-continuous, $\mu(A) = \mu(\text{cl}A)$ and $A^\varepsilon = (\text{cl}A)^\varepsilon$. $\triangleleft$

Lemma 3. *Let A be a subset of X. The set of real numbers $\varepsilon > 0$ for which A^ε is not μ-continuous is a countable set.*

Proof. In order to prove this, take a set A in X. Notice that for $\varepsilon > 0$, the set ∂A^ε is a subset of the set $\{x \in X \mid d(x, A) = \varepsilon\}$. So, the intersection of ∂A^ε and ∂A^δ is empty as soon as ε is not equal to δ.

We have to show that there are at most countably many numbers $\varepsilon > 0$ for which $\mu(\partial A^\varepsilon) > 0$. To this end, let n be a natural number. Suppose that there are positive numbers $\varepsilon_1, \ldots, \varepsilon_{n+2}$ such that $\varepsilon_k \neq \varepsilon_l$ whenever $k \neq l$ and moreover

$$\frac{1}{n+1} \leq \mu(\partial A^{\varepsilon_k}) \quad \text{for all } k = 1, \ldots, n+2.$$

Then by additivity of μ and the fact mentioned above that the sets $\partial A^{\varepsilon_k}$ are mutually disjoint,

$$1 \geq \mu(\bigcup_{k=1}^{n+2} \partial A^{\varepsilon_k}) = \sum_{k=1}^{n+2} \mu(\partial A^{\varepsilon_k}) \geq \sum_{k=1}^{n+2} \frac{1}{n+1} = \frac{n+2}{n+1} > 1$$

which is a contradiction. Consequently, there is a finite number of numbers $\varepsilon > 0$ with $\mu(\partial A^\varepsilon) \geq \frac{1}{n+1}$, and therefore countably many numbers $\varepsilon > 0$ with $\mu(\partial A^\varepsilon) > 0$. $\triangleleft$

Finally in this section we will prove a technical statement concerning the link between integrals over λ_i and those over beliefs $\mu_{i\tau}(hi\tau)$.

Lemma 4. *Let ϕ be a bounded and $\mathcal{B}_{\tau+1}(H_i)$-measurable function. Then we have*

$$\int_{H_i} \phi(h_i) d\lambda_i$$
$$= \int_{H_{i\tau}} \int_{X_i} \phi(h_{i\tau}, x_{i\tau+1}) \int_{\Theta_i} g_i(x_{i\tau+1}\theta_i \mid p_{i\tau}) d\mu_{i\tau}(h_{i\tau}) d\nu_i d\lambda_{i\tau}.$$

Proof. Let $D \times D_{\tau+1} \times \prod_{t=\tau+1}^{\infty} X_i$ be a Borel set in $\mathcal{B}_{\tau+1}(H_i)$. Then

$$\begin{aligned} & \int_{H_i} \mathbf{1}_{D\times D_{\tau+1}\times\prod_{t=\tau+2}^{\infty} X_i} d\lambda_i = \int_{H_{i,\tau+1}} \mathbf{1}_{D\times D_{\tau+1}} d\lambda_{i\tau+1} \\ = \quad & \lambda_{i\tau+1}(D \times D_\tau) = \int_{H_{i\tau}} \mathbf{1}_D \cdot \gamma_{i\tau+1}(h_{i\tau})(D_{\tau+1}) d\lambda_{i\tau} \\ = \quad & \int_{H_{i\tau}} \int_{X_i} \mathbf{1}_{D\times D_{\tau+1}} \int_{\Theta_i} g_i(x_{i\tau+1}, \theta_i \mid p_{i\tau}) d\mu_{i\tau}(h_{i\tau}) d\nu_i d\lambda_{i\tau}. \end{aligned}$$

The same equality now easily follows for an arbitrary integrable function. ◃

APPENDIX B. WEAK CONVERGENCE

In the text we discuss the convergence of beliefs over time. The type of convergence we use is commonly known as weak convergence on the set $\mathbb{P}(\Theta)$ of probability measures on Θ. That is, we apply the following concept. Let $C(\Theta)$ be the collection of continuous functions $f\colon \Theta \to \mathbb{R}$. Note that each of these functions is bounded, since Θ is compact. With each f in $C(\Theta)$ and $\mu \in \mathbb{P}(\Theta)$ we can therefore associate a number

$$\int_\Theta f(\theta) d\mu,$$

the integral of f with respect to μ. We will use the following terminology.

Definition 10. A sequence $(\mu_n)_{n=1}^{\infty}$ of probability measures in $\mathbb{P}(\Theta)$ *converges weakly* to a probability measure μ_∞ in $\mathbb{P}(\Theta)$ if for each $f \in C(\Theta)$

$$\int_\Theta f(\theta) d\mu_n \to \int_\Theta f(\theta) d\mu_\infty.$$

Notice that this is just a definition. It is clear that there is a topology in which the above sequences do converge. What is not immediately clear is that there is a topology in which these are the *only* convergent sequences. Nevertheless, we will show that this is the case, and also provide a number of different descriptions of this topology.

TOPOLOGY

One helpful interpretation of this notion of convergence of measures is in terms of pointwise convergence of functionals [5]. Let $C(\Theta)^*$ be the collection of func-

[5] A functional is a linear function from some vector space to the real numbers.

tionals on Θ that are continuous with respect to the max-norm on $C(\Theta)$. The max-norm $\|f\|_\infty$ of a function f in $C(\Theta)$ is defined as the real number

$$\|f\|_\infty := \max\{|f(\theta)| \mid \theta \in \Theta\}.$$

The collection $C(\Theta)^*$ is called the (first) dual space of $C(\Theta)$. We can say that a sequence $(I_n)_{n=1}^\infty$ in $C^*(\Theta)$ converges *pointwise* to $I \in C^*(\Theta)$ if for all points f in the domain $C(\Theta)$

$$I_n(f) \to I(f).$$

Now let μ be a probability measure in $\mathbb{P}(\Theta)$. With this probability measure we can associate a functional $I(\mu)$ in $C(\Theta)^*$ by

$$I(\mu)(f) := \int_\Theta f(\theta) d\mu.$$

Then, it is easily seen that $(\mu_n)_{n=1}^\infty$ converges weakly to μ_∞ if and only if $(I(\mu_n))_{n=1}^\infty$ converges pointwise to $I(\mu_\infty)$. Thus weak convergence is linked to the product topology on $C(\Theta)^*$.

Weak convergence of a sequence of probability measures is also related to the concept of topological convergence. In order to see this connection, take a sequence $x_0, x_1, x_2, \ldots$ of elements of a topological space (X, τ). We say that the sequence converges to x *in topology* τ if for every set $U \in \tau$ with $x \in U$ there exists an $N \in \mathbb{N}$ such that $x_n \in U$ for all $n \geq N$.

Now there is a topology on $\mathbb{P}(\Theta)$ such that the converging sequences according to this topology coincide with the weakly converging sequences. This topology is called the *weak topology* on $\mathbb{P}(\Theta)$ and it is denoted by $\mathcal{W}$.

$\mathcal{W}$ is defined as the topology generated by the collection B of sets $B \subset \mathbb{P}(\Theta)$ for which there is a probability measure μ in $\mathbb{P}(\Theta)$ and a sequence $f_1, \ldots, f_n$ of continuous functions on Θ as well as a sequence $\varepsilon_1, \ldots, \varepsilon_n$ of positive numbers in $\mathbb{R}$ such that

$$B = \left\{ \nu \in \mathbb{P}(\Theta) \mid \left| \int_\Theta f_k(\theta) d\mu - \int_\Theta f_k(\theta) d\nu \right| < \varepsilon_k \text{ for all } k = 1, \ldots, n \right\}.$$

It is elementary to check that B is indeed a basis and that convergence in the topology $\mathcal{W}$ generated by it coincides with weak convergence.

In terms of applicability a more convenient basis for $\mathcal{W}$ is the collection C of subsets C of Θ for which there is a probability measure μ on Θ, a sequence $A_1, \ldots, A_n$ of μ-continuous Borel sets and a sequence $\varepsilon_1, \ldots, \varepsilon_n$ of positive numbers such that

$$C = \{\nu \mid |\mu(A_k) - \nu(A_k)| < \varepsilon_k \quad \text{for all } k = 1, \ldots, n\}.$$

At least it is clear that C is indeed a basis and therefore generates some topology. Before we show that the topology generated by C is indeed $\mathcal{W}$ we will first show a result that is somewhat stronger than strictly necessary in the proof. We need it in its full strength later though.

Lemma 5. *Let μ be a probability measure on Θ and let f be a continuous function on Θ. Further suppose that we have a closed μ-continuous set F and a real number $\varepsilon > 0$. Then the set B of probability measures ν for which*

$$\left| \int_F f(\theta) d\nu - \int_F f(\theta) d\mu \right| < \varepsilon$$

contains an element C of C with μ in C.

Proof. We may assume w.l.o.g. that $0 \leq f(\theta) \leq 1$. Take a natural number s in $\mathbb{N}$. Then, since f is continuous and F is a closed μ-continuous set, using Lemma 3 we can construct closed μ-continuous sets $G_{0s} \supset G_{1s} \supset \cdots \supset G_{ss}$ in F such that

(i) $f(\theta) \geq \frac{k}{s} - \frac{1}{s^2}$ for all $\theta \in G_{ks}$ and
(ii) $f(\theta) < \frac{k}{s}$ for all $\theta \in F \setminus G_{ks}$ [6].

Write $\quad R_{ks} := G_{ks} \setminus G_{k+1,s}$ for $k = 0, \ldots, s-1$ and $R_{ss} := G_{ss}$.

Then $$\sum_{k=0}^{s} (\frac{k}{s} - \frac{1}{s^2}) \mu(R_{ks}) \leq \int_F f(\theta) d\mu \leq \sum_{k=0}^{s} \frac{k+1}{s} \mu(R_{ks}),$$

[6] Note that this condition implies $G_{0s} = F$.

while

$$\begin{aligned}
&\left|\sum_{k=0}^{s}(\frac{k}{s}-\frac{1}{s^2})\mu(R_{ks})-\sum_{k=0}^{s}\frac{k+1}{s}\mu(R_{ks})\right| \\
= &\sum_{k=0}^{s}(\frac{k+1}{s}-\frac{k}{s}+\frac{1}{s^2})\mu(R_{ks})=(\frac{1}{s}+\frac{1}{s^2})\sum_{k=0}^{s}\mu(R_{ks}) \\
= &(\frac{1}{s}+\frac{1}{s^2})\mu(F).
\end{aligned}$$

Now take a natural number t in $\mathbb{N}$ such that $(\frac{1}{t}+\frac{1}{t^2})\mu(F)<\frac{1}{2}\varepsilon$. Then $R_{ot},\ldots,R_{tt}$ is a finite number of μ-continuous Borel sets. So, the collection of probability measures ν for which for all R_{kt}

$$|\nu(R_{kt})-\mu(R_{kt})|<\frac{\varepsilon}{2(t+1)(t+2)}$$

is an element of C.

Now take such a ν. We will show that it is an element of the set B specified in the Lemma as well. To this end, notice that

$$\begin{aligned}
&\left|\sum_{k=0}^{t}(\frac{k+1}{t})\nu(R_{kt})-\sum_{k=0}^{t}\frac{k+1}{t}\mu(R_{kt})\right| \\
\leq &\sum_{k=0}^{t}(\frac{k+1}{t})\,|\nu(R_{kt})-\mu(R_{kt})|\leq\sum_{k=0}^{t}(\frac{k+1}{t})\frac{\varepsilon}{2(t+1)(t+2)} \\
= &\frac{\varepsilon}{2t(t+1)(t+2)}\sum_{k=0}^{t}(k+1)=\frac{\varepsilon}{2t(t+1)(t+2)}\frac{1}{2}(t+1)(t+2) \\
= &\frac{\varepsilon}{4t}\leq\frac{1}{4}\varepsilon.
\end{aligned}$$

A similar argument holds for the lower bounds on the respective integrals, so

$$\left|\int_F f(\theta)d\nu-\int_F f(\theta)d\mu\right|\leq\frac{3}{4}\varepsilon<\varepsilon. \qquad \triangleleft$$

This enables us to show

Lemma 6. *The topology generated by* C *coincides with* $\mathcal{W}$.

Proof. A. First we will show that the topology generated by C is a subset of $\mathcal{W}$. To see this, take a probability measure μ, a μ-continuous set Borel set A

and a real number $\eta > 0$. It is sufficient to show that the set

$$C := \{\nu \mid |\mu(A) - \nu(A)| < \eta\}$$

contains an element of the above basis $\mathtt{B}$ of $\mathcal{W}$.

In order to show that, notice that A is a μ-continuous Borel set. So, by Corollary 2, we know that there is an $\varepsilon > 0$ such that $\mu(A^\varepsilon) - \mu(A) < \frac{1}{2}\eta$. Further, the function $f: \Theta \to \mathbb{R}$ defined by

$$f(\theta) := (1 - \varepsilon^{-1} d(\theta, A)) \vee 0$$

is continuous. So, the collection of probability measures ν such that

$$\left| \int_\Theta f(\theta) d\mu - \int_\Theta f(\theta) d\nu \right| < \frac{1}{2}\eta$$

is an element of $\mathtt{B}$. We will show that it is a subset of C. To this end, notice that $\mathbf{1}_A \le f \le \mathbf{1}_{A^\varepsilon}$ on Θ. Using this fact, together with the above inequalities we get

$$\begin{aligned} \nu(A) &\le \int_\Theta f(\theta) d\nu < \int_\Theta f(\theta) d\mu + \frac{1}{2}\eta \\ &\le \mu(A^\varepsilon) + \frac{1}{2}\eta \le \mu(A) + \frac{1}{2}\eta + \frac{1}{2}\eta = \mu(A) + \eta. \end{aligned}$$

The other inequality follows from the same line of reasoning applied to the μ-continuous set Borel set $\Theta \setminus A$.

B. Conversely, suppose that we have a set of the form

$$\{\nu \mid \left| \int_\Theta f(\theta) d\mu - \int_\Theta f(\theta) d\nu \right| < \varepsilon\}$$

for some continuous f and $\varepsilon > 0$. Then, since Θ is a closed μ-continuous set, it must contain an element of $\mathtt{C}$ by Lemma 5. $\triangleleft$

Next we will show that the weak topology also coincides with the topology induced by the following distance function on $\mathbb{P}(\Theta)$.

Let μ and ν be two elements of $\mathbb{P}(\Theta)$. Then the *Prohorov distance* $\rho(\mu, \nu)$ is defined as the infimum over those real numbers $\varepsilon > 0$ for which every Borel set

A in Θ satisfies both

$$\mu(A) \leq \nu(A_\varepsilon) + \varepsilon \quad \text{and} \quad \nu(A) \leq \mu(A_\varepsilon) + \varepsilon.$$

First we will establish

Theorem 10. *The Prohorov distance ρ is a metric on the set $\mathbb{P}(\Theta)$.*

Proof. The only condition whose proof is not straightforward is the assertion that $\rho(\mu, \nu) = 0$ implies $\mu = \nu$.

So, assume that $\rho(\mu, \nu) = 0$. Take a closed set F in Θ. We will show that $\mu(F) = \nu(F)$. To this end, take a positive number $\eta > 0$. By Lemma 2 we know that $\nu(F^\varepsilon) - \nu(F) < \eta$ for all sufficiently small $\varepsilon > 0$. Furthermore, since the Prohorov distance between μ and ν equals zero, we also know that

$$\mu(F) \leq \nu(F_\varepsilon) + \varepsilon$$

for all these $\varepsilon > 0$. Together this yields

$$\mu(F) \leq \nu(F) + \eta + \varepsilon$$

for all sufficiently small $\varepsilon > 0$. Since $\eta > 0$ was also arbitrary we find that $\mu(F) \leq \nu(F)$. The converse inequality follows by symmetry.

So now we now that μ and ν coincide on closed sets. However, since μ and ν are probability measures, this immediately implies that they coincide on open sets as well. Then though they must coincide on all Borel sets by the regularity of both measures. $\triangleleft$

This implies that the Prohorov distance induces a Hausdorff topology, one that we will call the Prohorov topology for the moment. In order to show that it coincides with the weak topology we need

Lemma 7. *Let μ be a probability measure on Θ and let $\varepsilon > 0$ be a real number. There exists a finite partition $\mathcal{A}$ of Θ such that each A in $\mathcal{A}$ is a μ-continuous Borel set and* $\operatorname{diam}(A) \leq \varepsilon$*. Additionally, $\mathcal{A}$ can be constructed in such a way that each A in $\mathcal{A}$ has a non-empty interior.*

Proof. Notice that, by Lemma 3, we can choose for each θ in Θ a positive $\delta(\theta) < \frac{1}{2}\varepsilon$ such that $\theta_{\delta(\theta)}$ is μ-continuous. Since the collection of these sets cover the compact set Θ we can find $\theta(1), \ldots, \theta(n)$ such that the finite collection of open sets $B_k := \theta(k)_{\delta(\theta(k))}$ still covers Θ. Let $\mathcal{A}$ be the collection of non-empty sets of the form

$$\bigcap_{k \in K} B_k \cap \bigcap_{k \notin K} \Theta \setminus B_k$$

for some subset K of $\{1, \ldots, n\}$. This is clearly a partition of Θ. Furthermore, since each element A of $\mathcal{A}$ is a finite intersection of μ-continuous Borel sets B_k and their complements, it is easy to see that each element of $\mathcal{A}$ is also a μ-continuous Borel set. Finally, since $B_1, \ldots, B_n$ covers Θ, each A in $\mathcal{A}$ must be contained in at least one B_k by non-emptiness of A.

The additional requirement of non-empty interior can be guaranteed as well. The proof of this is in two parts. Let

$$\mathcal{A} = \{A_1, \ldots, A_n\}$$

be a finite partition of Θ such that $\mathrm{diam}(A) \leq \varepsilon$ and moreover each A_k is the intersection of an open set U_k and a closed set G_k. Notice that the above partition indeed has these properties. We will show how to construct a partition whose elements have non-empty interior.

Let $\mathcal{N}$ be the (possibly empty) collection of sets A in $\mathcal{A}$ whose interior is not empty. Let $A_1, \ldots, A_m$ be an enumeration of $\mathcal{N}$. Define

$$B_1 := A_1 \cup \bigcup_{A \notin \mathcal{N}} \mathrm{cl}(A_1) \cap A$$

and iteratively for each $2 \leq k \leq m$

$$B_k := \Big(A_k \cup \bigcup_{A \notin \mathcal{N}} \mathrm{cl}(A_k) \cap A\Big) \setminus \bigcup_{i \leq k-1} B_i.$$

We will show that $B_1, \ldots, B_m$ satisfies all our requirements. It is immediate that it is a sequence of mutually disjoint Borel sets. Furthermore, since $A_k \subset B_k \subset \mathrm{cl}(A_k)$ it is also immediate that each B_k has non-empty interior and diameter less than or equal to ε. So, we only have to show that $B_1, \ldots, B_m$ covers Θ.

Suppose that there exists an element θ in Θ that is not covered by any B_k. Then it is certainly not an element of any A in $\mathcal{N}$. So, since $\mathcal{A}$ covers Θ, it must be an element of some $A^* \notin \mathcal{N}$. Now suppose that it is also an element of the closure of some A_k in $\mathcal{N}$. Then it is also an element of

$$A_k \cup \bigcup_{A \notin \mathcal{N}} \mathrm{cl}(A_k) \cap A$$

since it is specifically an element of $\mathrm{cl}(A_k) \cap A^*$. This though implies that it is either an element of B_k or an element of $\bigcup_{i \le k-1} B_i$. Both cases contradict the assumption that x is not covered by any B_k. Hence, θ is not an element of the closure of any A in $\mathcal{N}$.

Since Θ is compact and $\mathcal{N}$ is finite, this implies that there is an $\varepsilon > 0$ such that θ_ε does not intersect any A in $\mathcal{N}$. So, θ_ε must be covered by the elements in $\mathcal{A} \setminus \mathcal{N}$. We will derive a contradiction. Let $A_1, \ldots, A_t$ be an enumeration of $\mathcal{A} \setminus \mathcal{N}$. The claim is that there is at least one A_i that is dense some non-empty open subset of θ_ε. Suppose not. Then in particular A_1 is not dense on any non-empty open subset of x_ε. So, there is a $\theta(1)$ in θ_ε and an $\varepsilon(1) > 0$ such that $\theta(1)_{\varepsilon(1)}$ has an empty intersection with A_1. This implies that $\theta(1)_{\varepsilon(1)}$ is covered by $A_2, \ldots, A_n$. Iteration of this argument eventually yields an open set $\theta(t+1)_{\varepsilon(t+1)}$ that has empty intersection with all A_k in $\mathcal{A} \setminus \mathcal{N}$. However, since $\theta(t+1)_{\varepsilon(t+1)}$ is a subset of θ_ε this means that it has an empty intersection with every A in $\mathcal{A}$ which contradicts the assumption that $\mathcal{A}$ covers Θ.

So we can take a non-empty open set V and a set $A_k = U_k \cap G_k$ in $\mathcal{A}$ whose interior is empty such that A_k is dense on V_k. Then it is certainly true that G_k is dense on the non-empty (!) open set $V \cap U_k$. This however implies that $V \cap U_k$ is a subset of G_k, since G_k is closed. So, the non-empty open set $V \cap U_k$ is a subset of $U_k \cap G_k = A_k$ and A_k has a non-empty interior. Contradiction. Hence, θ is an element of some B_k and $B_1, \ldots, B_n$ is a cover of Θ. $\triangleleft$

Theorem 11. *The Prohorov topology coincides with the weak topology $\mathcal{W}$.*

Proof. A. First we will show that the weak topology is a subset of the Prohorov topology. To this end, let μ be a probability measure on Θ. Further, let A be a μ-continuous Borel set in Θ and let $\varepsilon > 0$. It is sufficient to show that

$$C := \{\nu \in \mathbb{P}(\Theta) \mid |\mu(A) - \nu(A)| < \eta\}$$

is an element of the Prohorov topology by Lemma 6. To do that, it is even sufficient to show that there exists a real number $\delta > 0$ such that the collection of probability measures ν with $\rho(\mu, \nu) < \delta$ is a subset of C.

To this end, notice that A is assumed to be a μ-continuous Borel set. So, by Corollary 2, we know that there is an $\varepsilon > 0$ such that $\mu(A^\varepsilon) < \mu(A) + \frac{1}{2}\eta$. We can even take ε such that $\varepsilon < \frac{1}{2}\eta$. Then, for ν with $\rho(\mu, \nu) < \varepsilon$,

$$\nu(A) \leq \mu(A^\varepsilon) + \varepsilon < \mu(A) + \frac{1}{2}\eta + \frac{1}{2}\eta = \mu(A) + \eta.$$

In order to get the converse inequality $\nu(A) > \mu(A) - \eta$ we can simply apply the above line of reasoning to the μ-continuous Borel set $X \setminus A$.

B. Secondly we will show that the Prohorov topology is a subset of the weak topology. To this end, take a probability measure μ on Θ and a real number $\varepsilon > 0$. By Lemma 6 it is sufficient to show that the collection of probability measures ν with Prohorov distance less than ε to μ contains an element of the basis C of $\mathcal{W}$ described above.

Take a partition $\mathcal{A}$ of Θ as in Lemma 7. Then it is clear that the set

$$C := \{\nu \mid |\nu(A) - \mu(A)| < |\mathcal{A}|^{-1}\varepsilon \quad \text{for all } A \in \mathcal{A}\}$$

is an element of C. So we only need to show that all elements of C have Prohorov distance less than ε to μ.

In order to do this, take an element ν of C. Furthermore, let B be a Borel set in Θ. Let $\mathcal{S}$ be the set of elements A of $\mathcal{A}$ for which $B \cap A$ is not empty. Then, since $\mathcal{A}$ is a cover of Θ, B is a subset of $S := \bigcup_{A \in \mathcal{S}} A$. Moreover, since $\operatorname{diam}(A) < \varepsilon$ for all A, S is a subset of B_ε. Therefore we have

$$\mu(B) \leq \mu(S) = \sum_{A \in \mathcal{S}} \mu(A) < \sum_{A \in \mathcal{S}} (\nu(A) + |\mathcal{A}|^{-1}\varepsilon) < \nu(S) + \varepsilon \leq \nu(B_\varepsilon) + \varepsilon.$$

Similarly $$\nu(B) \leq \nu(S) < \mu(S) + \varepsilon \leq \mu(B_\varepsilon) + \varepsilon$$

which completes the proof. ◁

COMPACTNESS

We need to establish one more topological feature of $\mathbb{P}(\Theta)$, its compactness. We will provide a complete and detailed proof along the lines of the direct Theorem of Prohorov. However, we will bypass the embedding Theorem of Urysohn. First we need some general theory. In this section (K, d) will be an arbitrary complete metric space.

Definition 11. We say that K is *sequentially compact* if every sequence has a convergent subsequence.

Definition 12. Suppose we have a real number $\varepsilon > 0$. A *finite ε-cover* of K is a finite collection of open sets

$$x(1)_\varepsilon, \ldots, x(n)_\varepsilon$$

with center point x_k and radius ε that cover K.

Still under the assumption that K is complete and metric we have

Theorem 12. *The following three statements are equivalent.*

(1) *K is compact*

(2) *K is sequentially compact and separable*

(3) *For every real number $\varepsilon > 0$ there exists a finite ε-cover of K.*

Proof. We will show the implications in the cycle (1) → (3) → (2) → (1).

(1) → (3). Suppose that K is compact. Let $\varepsilon > 0$. Then the collection of open sets $x\varepsilon$ with x in K is obviously an open cover of K. Hence, by compactness, it has a finite subcover and this subcover obviously is a finite ε-cover of K.

(3) → (2). Suppose that (3) holds. Take a sequence $(x_n)_{n=1}^{\infty}$ in K. We have to show that this sequence has a convergent subsequence. To this end, take a natural number k. Then by assumption we can find points $y(k1), \ldots, y(ks(k))$

such that

$$y(k1)_{\frac{1}{k}}, \ldots, y(ks(k))_{\frac{1}{k}}$$

covers K. Now consider the following construction. Since

$$y(11)_1, \ldots, y(1s(1))_1$$

covers K, there must be a $y(1t(1))$ such that $y(1t(1))_1$ contains an infinite of number points x_n. Let $x_{\alpha(1)}$ be the first. Furthermore, switch to a subsequence that is completely contained in $y(1t(1))_1$. Since

$$y(21)_{\frac{1}{2}}, \ldots, y(2s(2))_{\frac{1}{2}}$$

covers K, there must be a $y(2t(2))$ such that $y(2t(2))_{\frac{1}{2}}$ contains an infinite number of points x_n. Let $x_{\alpha(2)}$ be the first one that has index $\alpha(2) > \alpha(1)$. Notice that we can do that, since there is an infinite number of points x_n that satisfy our conditions. Furthermore, switch to a subsequence that is completely contained in $y(2t(2))_{\frac{1}{2}}$. Et cetera.

Thus we find a subsequence $(x_{\alpha(n)})_{n=1}^{\infty}$ of $x_1, x_2, \ldots$ such that $x_{\alpha(k)}, x_{\alpha(k+1)}, \ldots$ is completely contained in $y(kt(k))_{\frac{1}{k}}$. This however means that this subsequence is Cauchy. Hence, since K is complete, it must be convergent.

Finally notice that the collection of points $y(ks)$ for k in $\mathbb{N}$ and $1 \leq s \leq s(k)$ is a countable set that is dense in K. Hence, K is separable as well.

$(2) \rightarrow (1)$. Suppose that K is separable and sequentially compact. Let $\mathcal{A}$ be some index set and let $(U_\alpha)_{\alpha \in \mathcal{A}}$ be an open cover of K. Suppose it does not have a finite subcover. We will derive a contradiction.

Since K is separable and metric, we know that there is a countable basis $\mathcal{B}$ that generates the topology on K. Let $B_1, B_2, \ldots$ be an enumeration of those elements of $\mathcal{B}$ that are contained in some U_α. Since $\mathcal{B}$ is a basis, it is clear that the above sequence also covers K. Furthermore it is clear that does not have a finite subcover, since a finite subcover of $B_1, B_2, \ldots$ easily translates to a finite subcover of $(U_\alpha)_{\alpha \in \mathcal{A}}$.

Now consider the following construction. Take a point x_1 in $B_{k(1)} := B_1$. Now notice that $B_{k(1)}$ does not cover K. Therefore the minimal natural number k for which B_k is not a subset of $B_{k(1)}$ exists. Denote this number by $k(2)$ and take a point

$$x_2 \in B_{k(2)} \setminus B_{k(1)}.$$

Now $B_{k(1)}, B_{k(2)}$ does not cover K either. So the minimal number k for which B_k is not a subset of $B_{k(1)} \cup B_{k(2)}$ exists as well. Denote this number by $k(3)$. Automatically $k(3) > k(2)$. Take a point

$$x_3 \in B_{k(3)} \setminus B_{k(1)} \cup B_{k(2)}.$$

Et cetera. Thus we get a sequence $x_1, x_2, \ldots$ of points in K. By assumption this sequence has a subsequence $x_{\alpha(1)}, x_{\alpha(2)}, \ldots$ that converges to some point, say x, in K. Now, since $B_1, B_2, \ldots$ covers K, we know that x is an element of some B_m. Furthermore, B_m must be a subset of $\bigcup_{n=1}^{m} B_{k(n)}$ by construction. This though contradicts the fact that a tail of the sequence $x_{\alpha(1)}, x_{\alpha(2)}, \ldots$ is not contained in this union by construction. $\triangleleft$

Switching back to the original setting, consider the metric space $\mathbb{P}(\Theta)$ equipped with the Prohorov distance ρ. We will show its (sequential) compactness by showing that it has a finite ε-cover for each $\varepsilon > 0$. First of all we have

Theorem 13. *The metric space $(\mathbb{P}(\Theta), \rho)$ is complete.*

Proof. Let $\mu_1, \mu_2, \ldots$ be a Cauchy sequence of probability measures on Θ. We will show that it converges to a probability measure μ on Θ. To this end, let $\mathcal{R}$ be the collection of Borel sets A in Θ for which the sequence

$$\mu_1(A), \mu_2(A), \ldots$$

converges. Define the function $\nu: \mathcal{R} \to \mathbb{R}$ by

$$\nu(A) := \lim_{n \to \infty} \mu_n(A)$$

for all A in $\mathcal{R}$. First in Part A we will show that $\mathcal{R}$ is a ring and that ν is a premeasure on $\mathcal{R}$. So, ν has a unique extension to the σ-algebra generated by

$\mathcal{R}$. Then in Parts B and C we will show that the σ-algebra generated by $\mathcal{R}$ must be equal to the Borel σ-algebra and the proof is complete.

A. Using the σ-additivity of the probability measures μ_n it is straightforward to check that $\mathcal{R}$ is a ring and that ν is a pre-measure on the ring $\mathcal{R}$.

B. Let x be an element of Θ. In this part we will show that the collection of numbers $\varepsilon > 0$ for which x^ε is not an element of $\mathcal{R}$ is countable [7]. To this end define the function $f_n: \mathbb{R} \to \mathbb{R}$ by

$$f_n(\varepsilon) := \mu_n(x^\varepsilon).$$

Write $D := \text{diam}(\Theta)$. Then

(1) $f_n(\varepsilon) = 1$ for all $\varepsilon > D$ and $f_n(\varepsilon) = 0$ for all $\varepsilon < 0$

(2) f_n is non-decreasing

(3) f_n is cadlag by Lemma 2 and (2).

Now let $q_1, q_2, \ldots$ be an enumeration of the rational numbers. Since $f_1, f_2, \ldots$ is a bounded sequence, there is a subsequence $f_{11}, f_{12}, \ldots$ such that

$$f_{11}(q_1), f_{12}(q_1), \ldots$$

converges. Similarly we can take a subsequence $f_{21}, f_{22}, \ldots$ such that

$$f_{21}(q_2), f_{22}(q_2), \ldots$$

converges. Then it is not so hard to show that the subsequence $(f_{kk})_{k=1}^\infty$ converges for all q_l and we can define

$$f(q_l) := \lim_{k\to\infty} f_{kk}(q_l).$$

This is obviously a non-decreasing function on the rational numbers, so we can extend f to all real numbers by

$$f(r) := \inf\{f(q_l) \mid q_l \geq r\}.$$

[7] The argument used here is basically the proof of Helly's Theorem tailored to our special situation.

It is elementary to check that f satisfies (1) till (3).

Now take a real number r such that f is continuous in r. We want to show that

$$f(r) = \lim_{n\to\infty} f_n(r).$$

Let $\eta > 0$. First we will show that there exists an N such that

$$f_m(r) \leq f(r) + \eta$$

for all $m \geq N$. First of all, notice that we can take a rational number $q \geq r$ such that $f(q) \leq f(r) + \frac{\eta}{3}$. We can assume w.l.o.g. that $\delta := q - r < \frac{\eta}{3}$.

Next, having chosen $\delta = q - r$, we can choose a natural number N_1 such that $\rho(\mu_n, \mu_m) < \delta$ for all $m, n \geq N_1$ since the sequence $\mu_1, \mu_2, \ldots$ is Cauchy.

Further, write $f_{\alpha(n)} := f_{kk}$. Since $f_{\alpha(1)}(q), f_{\alpha(2)}(q), \ldots$ converges to $f(q)$ we can choose a natural number N_2 such that $f_{\alpha(n)}(q) < f(q) + \frac{\eta}{3}$ for all $n \geq N_2$.

Now take n such that $n \geq N_1$ and $n \geq N_2$. Then, since $\alpha(n) \geq n$, for all $m \geq n$,

$$\begin{aligned} f_m(r) &= \mu_m(x^r) \leq \mu_{\alpha(n)}((x^r)_\delta) + \delta \leq \mu_{\alpha(n)}(x^{r+\delta}) + \delta \\ &= f_{\alpha(n)}(q) + \delta \leq f(q) + \tfrac{\eta}{3} + \delta \leq f(r) + \tfrac{\eta}{3} + \tfrac{\eta}{3} + \delta \leq f(r) + \eta. \end{aligned}$$

So, if $f_1(r), f_2(r), \ldots$ does not converge to $f(r)$, then there is a real number $\kappa > 0$ and a subsequence $f_{\beta(1)}(r), f_{\beta(2)}(r), \ldots$ converging to $f(r) - \kappa$. Take L such that $f_{\beta(l)}(r) \leq f(r) - \frac{4}{5}\kappa$ for all $\beta(l) \geq L$.

First of all, since f is assumed to be continuous in r, we can choose a rational number $q < r$ such that $f(q) \geq f(r) - \frac{1}{5}\kappa$. We may assume that $\delta := r - q < \frac{1}{5}\kappa$.

Choose N such that $\rho(\mu_m, \mu_n) < \delta$ for all $m, n \geq N$. Since $f_{\alpha(1)}(q), f_{\alpha(2)}(q), \ldots$ converges to $f(q)$ we know that we can take an $\alpha(n) \geq N$ such that $f_{\alpha(n)}(q) \geq f(q) - \frac{1}{5}\kappa$. So, on one hand,

$$f_{\alpha(n)}(q) \geq f(q) - \tfrac{1}{5}\kappa \geq f(r) - \tfrac{2}{5}\kappa.$$

On the other hand, take an l such that $\beta(l) \geq N$ and $\beta(l) \geq L$. Then

$$\begin{aligned} f_{\alpha(n)}(q) &= \mu_{\alpha(n)}(x^q) \leq \mu_{\beta(l)}((x^q)_\delta) + \delta \leq f_{\beta(l)}(q + \delta) + \delta \\ &\leq f_{\beta(l)}(r) + \tfrac{1}{5}\kappa \leq f(r) - \tfrac{4}{5}\kappa + \tfrac{1}{5}\kappa = f(r) - \tfrac{3}{5}\kappa \end{aligned}$$

and we have a contradiction. Now since f is non-decreasing, f only has a countable number of discontinuity points and the proof of Part B is complete.

C. We will show now that the σ-algebra generated by $\mathcal{R}$ equals the Borel σ-algebra. Since $\mathcal{R}$ is a subset of the Borel σ-algebra by definition, it suffices to show that all closed sets are included in the σ-algebra generated by $\mathcal{R}$.

To this end, let F be a closed subset of Θ. Take a natural number n. By Part B we can choose for every θ in Θ a real number $0 < \varepsilon(\theta) < \frac{1}{n}$ such that $\theta^{\varepsilon(\theta)}$ is an element of $\mathcal{R}$. Now we can choose a finite cover

$$\theta_1^{\varepsilon(\theta_1)}, \ldots, \theta_n^{\varepsilon(\theta_n)}$$

of F by compactness of F. Then it is clear that

$$F \subset \bigcup_{k=1}^{n} \theta_k^{\varepsilon(\theta_k)} \subset F_{\frac{1}{n}}$$

while the middle set is an element of $\mathcal{R}$ since it is a finite union of elements of $\mathcal{R}$. Hence, F can be written as a countable intersection of elements of $\mathcal{R}$ and must therefore be an element of the σ-algebra generated by $\mathcal{R}$. This concludes the proof. $\triangleleft$

Secondly,

Theorem 14. *For every $\varepsilon > 0$, $\mathbb{P}(\Theta)$ has a finite ε-cover.*

Proof. Take a real number $\varepsilon > 0$. Take a partition

$$\mathcal{A} = \{A_1, \ldots, A_n\}$$

as in Lemma 7. Take points x_k in $\text{int}(A_k)$ and a natural number T such that $T^{-1} < |\mathcal{A}|^{-1}\varepsilon$. These remain fixed throughout the proof.

Let $\delta(x_k)$ denote the Dirac measure on x_k. Let for each k a natural number $0 \leq t(k) \leq T$ be specified such that these numbers sum up to T. Then

$$\nu := \frac{1}{T} \sum_{k=1}^{n} t(k)\delta(x_k)$$

is a probability measure. Furthermore, each A_k is ν-continuous, since ν is a convex combination of Dirac measures $\delta(x_m)$ that are constructed in such a way that all A_k are $\delta(x_m)$-continuous. Therefore,

$$C(\nu) := \{\mu \mid |\mu(A) - \nu(A)| < |\mathcal{A}|^{-1}\varepsilon \quad \text{for all } A \in \mathcal{A}\}$$

is an element of the basis $\mathbf{C}$ and is therefore included in the set of probability measures μ that have Prohorov distance less than ε to ν by part B of the proof of Theorem 11. It is also clear that there is only a finite number of such sets, since the amount of probability measures of the form

$$\frac{1}{T}\sum_{k=1}^{n} t(k)\delta(x_k)$$

is finite. We will show that the collection of these sets covers $\mathbb{P}(\Theta)$.

Take a probability measure μ on Θ. Now select for each $1 \leq k \leq n$ a natural number $0 \leq s(k) \leq T$ such that

$$\frac{s(k)}{T} \leq \mu(A_k) < \frac{s(k)+1}{T}.$$

Now the numbers $s(k)$ need not add up to T, but their sum is certainly less than or equal to T. Moreover, it is easy to select numbers $t(k) \in \{s(k), s(k)+1\}$ in such a way that the numbers $t(k)$ do add up to T, the only restriction here being that $\mathcal{A}$ has at least two elements. Finally, it is elementary to check that for ν defined by

$$\nu := \frac{1}{T}\sum_{k=1}^{n} t(k)\delta(x_k)$$

the probability measure μ is an element of $C(\nu)$. $\triangleleft$

Now we have developed enough equipment to prove

Theorem 15. $\mathbb{P}(\Theta)$ *is (sequentially) compact w.r.t. the weak topology.*

Proof. By Theorem 13 we know that $\mathbb{P}(\Theta)$ is complete with respect to the Prohorov distance. By Theorem 14 we know that it has a finite ε-cover for every $\varepsilon > 0$. Hence, by Theorem 12, it is also (sequentially) compact. $\triangleleft$

SEQUENTIAL THEOREMS

There are also several ways to check whether or not a sequence $(\mu_n)_{n=1}^{\infty}$ of probability measures converges weakly to some limit μ_∞ without directly using the topological framework. In this section we will state some of them.

Theorem 16. (Portmanteau) *Let $\mu_\infty, \mu_1, \mu_2, \ldots$ be probability measures on Θ. Then the following statements are equivalent.*
(1) *$\mu_1, \mu_2, \ldots$ converges weakly to μ_∞*
(2) *$\lim_{n\to\infty} \mu_n(A) = \mu_\infty(A)$ for all μ_∞-continuous Borel sets A*
(3) *$\lim_{n\to\infty} \mu_n(F) = \mu_\infty(F)$ for all closed μ_∞-continuity sets F in Θ.*

Proof. The equivalence of (1) and (2) follows from Lemma 6. We will show (2) $\to$(3) $\to$(2) to establish (3).

The implication from (2) to (3) is evident, so we only have to prove the converse implication. To that end, assume that we have a sequence $\mu_\infty, \mu_1, \mu_2, \ldots$ of probability measures with

$$\lim_{n\to\infty} \mu_n(F) = \mu_\infty(F)$$

for all closed μ_∞-continuity sets F. Let A be an arbitrary μ_∞-continuity set. We will show that

$$\lim_{n\to\infty} \mu_n(A) = \mu_\infty(A).$$

In order to do that, take an arbitrary real number $\varepsilon > 0$. We will show that there is a natural number N such that for all $n \geq N$

$$|\mu_n(A) - \mu_\infty(A)| < 2\varepsilon.$$

To this end notice that the closure $\mathrm{cl}A$ of A is a μ_∞-continuity set since $\partial \mathrm{cl}A = \partial A$. So by our assumption there is a natural number N_1 such that

$$|\mu_n(\mathrm{cl}A) - \mu_\infty(\mathrm{cl}A)| < \varepsilon$$

for all $n \geq N_1$. Furthermore, notice that the real number $\mu_\infty(\mathrm{cl}A)$ is equal to $\mu_\infty(A)$ since $\mu_\infty(\partial A) = 0$. Therefore it is sufficient to show that there is a natural number N_2 such that for all $n \geq N_2$

$$|\mu_n(\mathrm{cl}A) - \mu_n(A)| < \varepsilon.$$

To this end, notice that ∂A is closed and, since $\partial\partial A = \partial A$, a μ_∞-continuity set as well. So, by our assumption we know that

$$\lim_{n\to\infty} \mu_n(\partial A) = \mu_\infty(\partial A) = 0.$$

This implies that there is an N_2 such that for all $n \geq N_2$

$$|\mu_n(\partial A)| < \varepsilon.$$

Therefore, since $\mu_n(A \setminus \partial A) \leq \mu_n(A) \leq \mu_n(\mathrm{cl}A)$ by monotonicity of μ_n and $\mu_n(A \setminus \partial A) + \mu_n(\partial A) = \mu_n(\mathrm{cl}A)$ since $\mathrm{cl}A = A \cup \partial A$, we get that

$$|\mu_n(\mathrm{cl}A) - \mu_n(A)| < \varepsilon$$

for all $n \geq N_2$. This concludes the proof. ◁

Another variant we use in the paper is the following Lemma. Since it is an immediate consequence of Theorem 16 (3), its prrof is omitted. Suppose we have a subset $\mathcal{U}$ of the Borel σ-algebra $\mathcal{B}$ such that

(1) finite intersections of elements of $\mathcal{U}$ are also elements of $\mathcal{U}$, and

(2) each open set in Θ is the countable union of elements of $\mathcal{U}$.

Lemma 8. *A sequence $(\mu_n)_{n=1}^\infty$ in $\mathbb{P}(\Theta)$ converges weakly to a probability measure μ_∞ on Θ whenever $(\mu_n(U))_{n=1}^\infty$ converges to $\mu_\infty(U)$ for every element U of $\mathcal{U}$.*

APPENDIX C. MARTINGALE CONVERGENCE THEORY

In this section we work within a fixed probability space, denoted by $(\Omega, \aleph, \lambda)$. We will assume that Ω is a compact metric space. This is not strictly needed in the proofs, but it does make matters easier and it is the setup in which we will apply the results discussed here anyway. Further, $\aleph$ is assumed to be a σ-algebra on Ω and λ is a probability measure on $\aleph$.

CONDITIONAL EXPECTATION

In this section we will briefly discuss the theory concerning the existence and uniqueness of the conditional expected value of a random variable with respect to λ. The basic theorem is the Radon-Nikodym theorem. We will discuss that one first. Notice that we don't need the assumption of σ-finiteness since we assume that every measure is finite.

First we need some definitions. Assume for the moment that we have a σ-subalgebra $\mathcal{A}$ of $\aleph$ [8]. Further suppose that we have two measures μ and ν on $\mathcal{A}$.

Definition 13. The measure ν is said to be *absolutely continuous* with respect to μ if for every A in $\mathcal{A}$ with $\mu(A) = 0$ we have $\nu(A) = 0$.

Definition 14. An $\mathcal{A}$-measurable function $f: \Omega \to \mathbb{R}$ is a *density* of ν with respect to μ if for all A in $\mathcal{A}$

$$\nu(A) = \int_A f d\mu.$$

Theorem 17. (Radon-Nikodym) *Suppose that the measure ν is absolutely continuous with respect to μ. Then ν has a density with respect to μ. Moreover, if f and g are two such densities, then $f = g$ λ-almost surely* [9].

Using this result we can show the existence of the conditional expected value of a random variable [10].

Definition 15. A random variable X w.r.t. $\mathcal{A}$ is called *integrable* with respect to λ if

$$\int_\Omega |X(\omega)| d\lambda$$

is a real number.

Now let X be an integrable random variable w.r.t. λ.

Definition 16. An *expected value* of X *conditional on* $\mathcal{A}$ is an $\mathcal{A}$-measurable and integrable function f on Ω such that for all A in $\mathcal{A}$

$$\int_A f d\lambda = \int_A X d\lambda.$$

[8] A σ-subalgebra of $\aleph$ is a subset of $\aleph$ that is a σ-algebra.

[9] By this we mean that the collection of worlds ω where the equation is not true has λ-probability one.

[10] A random variable (w.r.t. $\mathcal{A}$) is simply a real-valued $\mathcal{A}$-measurable function on Ω.

EXISTENCE Existence of a conditional expected value of X on $\mathcal{A}$ can easily be derived from the Radon-Nikodym theorem. In order to do that, assume for the moment that X is non-negative. Then the formula

$$\nu(A) := \int_A X d\lambda \quad \text{for all } A \in \mathcal{A}$$

defines a measure on $\mathcal{A}$. Furthermore it is easy to check that this measure is absolutely continuous w.r.t. the restriction of λ to $\mathcal{A}$. So, according to the Radon-Nikodym theorem there exists an $\mathcal{A}$-measurable function f such that for every A in $\mathcal{A}$

$$\int_A f d\lambda = \nu(A) = \int_A X d\lambda.$$

So, this function f is indeed an expected value of X conditional on $\mathcal{A}$.

Now, for a general random variable, notice that the non-negative functions X_+ and X_- on Ω defined by

$$X_+(\omega) := \max\{X(\omega), 0\} \quad \text{and} \quad X_-(\omega) := \max\{-X(\omega), 0\}$$

are both random variables w.r.t. λ. So, there are expected values f_+ and f_- of X_+ and X_- resp. conditional on $\mathcal{A}$. It is now easy to check that $f := f_+ - f_-$ is an expected value of X conditional on $\mathcal{A}$. ◁

UNIQUENESS Now the second part of the Radon-Nikodym states that two different conditional expected values of X on $\mathcal{A}$ will be equal with probability one according to λ. This means that the collection of expected values of X conditional on $\mathcal{A}$ is an equivalence class of the equivalence relation $\sim$ on the collection of random variables on $\mathcal{A}$ defined by

$$f \sim g \quad \text{if and only if} \quad f = g \quad \lambda - \text{almost surely.}$$

This equivalence class is denoted by $\mathbb{E}(X \mid \mathcal{A})$. Any element of the class $\mathbb{E}(X \mid \mathcal{A})$ is called a *version* of $\mathbb{E}(X \mid \mathcal{A})$.

This class is obviously uniquely defined. Nevertheless we will slightly abuse notation and also use the symbol $\mathbb{E}(X \mid \mathcal{A})$ to indicate an element of this class.

In that sense the conditional expected value is defined only modulo sets having probability zero. ◁

The following simple observation will be used in the next section.

Lemma 9. *Let X and Y be two $\aleph$-measurable and integrable functions such that $X \leq Y$ with λ-probability one. Then*

$$\mathbb{E}(X \mid \mathcal{A}) \leq \mathbb{E}(Y \mid \mathcal{A})$$

with λ-probability one.

MARTINGALES Let $(\aleph_t)_{t=1}^{\infty}$ be a sequence of σ-subalgebras of $\aleph$, i.e. each σ-algebra $\aleph_t$ is a subset of $\aleph$. Such a sequence is said to *provide information* if $\aleph_t$ is a subset of $\aleph_{t'}$ for each $t' \geq t$. The expression "providing information" refers to the fact that in most applications the sequence of σ-algebras is generated by a sequence of partitions of Ω each partition reflecting the amount of information available at that time.

Definition 17. A sequence $(X_t)_{t=1}^{\infty}$ of random variables on Ω is said to *provide information* if each X_t is $\aleph_t$-measurable.

We will assume that such a sequence is uniformly bounded, i.e. there exists a number K such that for all t and ω

$$|X_t(\omega)| \leq K.$$

This requirement is of course only a technicality. We impose it because it makes life easier and because the condition is satisfied anyway in the application we use it for in the paper. Its main consequence is that each X_t is integrable w.r.t. λ and that the *expected value*

$$\mathbb{E}(|X_t|) := \int_{\Omega} |X_t(\omega)| d\lambda$$

of $|X_t|$ is also bounded by K.

Definition 18. A sequence $(X_t)_{t=1}^{\infty}$ that provides information is called a *submartingale* if

$$X_t \leq \mathbb{E}(X_{t+1} \mid \aleph_t)$$

for all t. If we even have equality the sequence is called a martingale.

Notice that the submartingale condition states that X_t is dominated by a version of $\mathbb{E}(X_{t+1} \mid \aleph_t)$. This means that the condition is equivalent to the requirement that

$$\int_A X_t d\lambda \leq \int_A X_{t+1} d\lambda$$

should hold for all A in $\aleph_t$. Similarly, being a martingale is equivalent with having equality in the displayed inequality. Of this formulation we will make particular use.

A martingale converges λ-almost surely. In other words, the probability that the sequence will keep changing, e.g. cycle, is zero. The remaining part of this section is devoted to a proof of this result.

So, let $(X_t, \aleph_t)_{t=1}^{\infty}$ be a submartingale. Let r be a real number. Define $Z_t \colon \Omega \to \mathbb{R}$ by

$$Z_t(\omega) := \max\{r, X_t(\omega)\}.$$

Lemma 10. *The sequence $(Z_t)_{t=1}^{\infty}$ is a submartingale.*

Proof. It is immediately clear that each Z_t is $\aleph_t$-measurable. Furthermore,

$$|Z_t(\omega)| = |\max\{r, X_t(\omega)\}| \leq \max\{|r|, |X_t(\omega)|\} \leq \max\{|r|, K\}$$

which implies that the sequence has a uniform upper bound. So we only need to check the submartingale condition. To this end notice that

$$X_{t+1} \leq Z_{t+1} \quad \text{and} \quad r \leq Z_{t+1}.$$

So, by Lemma 9,

$$\mathbb{E}(X_{t+1} \mid \aleph_t) \leq \mathbb{E}(Z_{t+1} \mid \aleph_t) \quad \text{and} \quad r = \mathbb{E}(r \mid \aleph_t) \leq \mathbb{E}(Z_{t+1} \mid \aleph_t)$$

with λ-probability one. Hence,

$$Z_t = \max\{r, X_t\} = \max\{r, \mathbb{E}(X_{t+1} \mid \aleph_t)\} \leq \mathbb{E}(Z_{t+1} \mid \aleph_t)$$

with λ-probability one. This concludes the proof. $\triangleleft$

From now on we will make the further assumption that we have a fixed world ω in Ω and a fixed natural number n. Only at the end of the proof these will become variable again.

Notice that $X_1(\omega), \ldots, X_n(\omega)$ is a sequence of real numbers. Now take two real numbers r and s with $r < s$. Define $T_0(\omega) := 0$, $T_1(\omega) := \min\{t \geq T_0 \mid x_t \leq r\}$ and recursively for $k = 2, 3, \ldots$

$$T_k(\omega) := \begin{cases} \min\{t > T_{k-1}(\omega) \mid x_t \leq r\} & \text{when } k \text{ is odd} \\ \min\{t \geq T_{k-1}(\omega) \mid x_t \geq s\} & \text{when } k \text{ is even} \end{cases}$$

until we are supposed to take the minimum over the empty set [11]. So, this yields an increasing sequence $T_0(\omega), T_1(\omega), \ldots, T_{K(n)(\omega)}(\omega)$ of natural numbers smaller than or equal to n.

With this sequence we can associate a sequence of indicator functions. Formally, for $1 \leq k \leq n$, let $I_k(\omega) \colon \mathbb{N} \to \{0, 1\}$ be defined by

$$I_k(\omega)(t) := \begin{cases} 1 & \text{when } T_{k-1}(\omega) < t \leq T_k(\omega) \\ 0 & \text{else.} \end{cases}$$

Strictly speaking, this is not a correct definition for $k > K(n)(\omega)$ since $T_k(\omega)$ is not defined for these values of k. We will interpret the definition for these cases though as if the corresponding function $I_k(\omega)$ is constantly equal to zero. We don't really need these functions $I_k(\omega)$ for values k larger than $K(n)(\omega)$, but they do keep notation simple in the proof. We will also use the shorthand notation

$$[I_k(t) = 1] := \{\omega \in \Omega \mid T_{k-1}(\omega) < t \leq T_k(\omega)\},$$

again with the convention that this is the empty set for values of k larger than $K(n)(\omega)$. Then we have

Lemma 11. *The set $[I_k(t) = 1]$ is $\aleph_{t-1}$-measurable.*

Proof. First notice that

$$[I_k(t) = 1] = \{\omega \mid T_{k-1}(\omega) < t \leq T_k(\omega)\} = \{\omega \mid T_{k-1}(\omega) < t\} \cap \{\omega \mid T_k(\omega) < t\}^c.$$

[11] Since $T_{k-1}(\omega) < T_k(\omega)$ it is easy to see that $\{t \mid T_k(\omega) < t \leq n\}$ has at most $n - k$ elements. From this it easily follows that we can perform the recursive step at most n times.

From this it easily follows that it is sufficient to show that the set

$$\{\omega \mid T_k(\omega) = u\}$$

is $\aleph_{t-1}$-measurable for each $0 \leq u \leq t-1$. This is what we will show now by induction to k.

Step 1. For $k = 0$. The set $\{\omega \mid T_0(\omega) = u\}$ is either equal to Ω (for $u = 0$) or to the empty set (for all other values of u). In both cases though it is clearly an element of $\aleph_{t-1}$.

Step $k+1$, in case $k+1$ is odd. Suppose we know that $\{\omega \mid T_k(\omega) = u\}$ is an element of $\aleph_{t-1}$ for all $0 \leq s \leq t-1$. Then

$$\{\omega \mid T_{k+1}(\omega) = u\} =$$
$$\bigcup_{v=0}^{u-1} \Big[\{\omega \mid T_k(\omega) = v\} \cap \{\omega \mid X_{v+1}(\omega) > r, \ldots, X_{u-1}(\omega) > r, X_u(\omega) \leq r\}\Big]$$

is $\aleph_{t-1}$-measurable by the induction hypothesis and the fact that $X_{v+1}, \ldots, X_u$ are $\aleph_{t-1}$-measurable. Obviously we can do something similar in case $k+1$ is even. $\triangleleft$

Let $U_n(\omega)$ be the largest even number k for which $T_k(\omega)$ exists. So,

$$U_n(\omega) := \max\{0 \leq k \leq K(n) \mid K \text{ is even}\}.$$

This number is called the number of upcrossings over (r, s). It counts the number of times the sequence goes from being less than or equal to r to being more than or equal to s. We have the following result.

Lemma 12. *Given the above setting, we have*

$$\mathbb{E}(U_n) \leq \frac{2}{s-r} \max\{|r|, K\}.$$

Proof. From Lemma 10 we already know that the sequence

$$Z_t := \max\{r, X_t\}$$

is a submartingale as well. Furthermore, it is easy to see that the random variables T_k, $K(n)$, $I_k(t)$ and U_n are identical for both $(X_t)_{t=1}^{\infty}$ and $(Z_t)_{t=1}^{\infty}$. So,

$$\begin{aligned} Z_n(\omega) - Z_1(\omega) &= \sum_{t=2}^{n} \Big(Z_t(\omega) - Z_{t-1}(\omega) \Big) \\ &= \sum_{t=2}^{n} \sum_{k=1}^{n} I_k(t)(\omega) \Big(Z_t(\omega) - Z_{t-1}(\omega) \Big) \end{aligned}$$

where the second equality follows from the observation that for each $2 \leq t \leq n$ exactly one element of the sequence $I_1(t)(\omega), \ldots, I_n(t)(\omega)$ will be equal to one, while the other elements are equal to zero. Now split the latter term, the double summation, into the two terms

$$E_n(\omega) := \sum_{t=2}^{n} \sum_{\substack{k=1 \\ k \text{ even}}}^{n} I_k(t)(\omega) \Big(Z_t(\omega) - Z_{t-1}(\omega) \Big)$$

and
$$O_n(\omega) := \sum_{t=2}^{n} \sum_{\substack{k=1 \\ k \text{ odd}}}^{n} I_k(t)(\omega) \Big(Z_t(\omega) - Z_{t-1}(\omega) \Big).$$

Notice that both E_n and O_n are λ-integrable since they are $\aleph$-measurable and bounded over Ω. In other words, they both have an expected value. We will treat the two terms separately for the moment and find lower bounds for their respective expected values.

A. Concerning the odd term O_n, notice that

$$\begin{aligned} \mathbb{E}(O_n) &= \int_{\Omega} O_n(\omega) d\lambda \\ &= \sum_{t=2}^{n} \sum_{\substack{k=1 \\ k \text{ odd}}}^{n} \int_{\Omega} I_k(t)(\omega) \Big(Z_t(\omega) - Z_{t-1}(\omega) \Big) d\lambda \\ &= \sum_{t=2}^{n} \sum_{\substack{k=1 \\ k \text{ odd}}}^{n} \Big(\int_{[I_k(t)=1]} Z_t(\omega) d\lambda - \int_{[I_k(t)=1]} Z_{t-1}(\omega) d\lambda \Big). \end{aligned}$$

However, since $(Z_t)_{t=1}^{\infty}$ is a submartingale, we get that

$$\int_{[I_k(t)=1]} Z_t(\omega) d\lambda - \int_{[I_k(t)=1]} Z_{t-1}(\omega) d\lambda \geq 0$$

for each t by Lemma 11. Hence, $\mathbb{E}(O_n) \geq 0$.

B. Concerning the even term E_n, notice that

$$\begin{aligned} E_n(\omega) &= \sum_{t=2}^{n} \sum_{\substack{k=1 \\ k \text{ even}}}^{n} I_k(t)(\omega)\Big(Z_t(\omega) - Z_{t-1}(\omega)\Big) \\ &= \sum_{\substack{k=1 \\ k \text{ even}}}^{n} \sum_{t=2}^{n} I_k(t)(\omega)\Big(Z_t(\omega) - Z_{t-1}(\omega)\Big) \\ &= \sum_{\substack{k=1 \\ k \text{ even}}}^{K(n)(\omega)} \sum_{t=2}^{n} I_k(t)(\omega)\Big(Z_t(\omega) - Z_{t-1}(\omega)\Big) \\ &= \sum_{\substack{k=1 \\ k \text{ even}}}^{K(n)(\omega)} \Big(Z_{T_k(\omega)}(\omega) - Z_{T_{k-1}(\omega)}(\omega)\Big) \geq (s-r)U_n(\omega). \end{aligned}$$

Hence, $\mathbb{E}(E_n) \geq (s-r)\mathbb{E}(U_n)$.

C. Combined, this yields

$$\begin{aligned} (s-r)\mathbb{E}(U_n) &\leq \mathbb{E}(E_n) + \mathbb{E}(O_n) = \mathbb{E}(Z_n - Z_1) \\ &\leq \mathbb{E}(|Z_n|) + \mathbb{E}(|Z_1|) \leq 2\max\{|r|, K\} \end{aligned}$$

which completes the proof. $\triangleleft$

We are now ready for the martingale convergence theorem. For each world ω in Ω for which the sequence $(X_t(\omega))_{t=1}^{\infty}$ converges, we define

$$X_\infty(\omega) := \lim_{t\to\infty} X_t(\omega).$$

Now we can prove

Theorem 18. *Let $(X_t, \aleph_t)_{t=1}^{\infty}$ be a martingale. Then X_∞ exists λ-almost surely.*

Proof. Suppose not. Let X_* be the $\aleph$-measurable function defined by

$$X_*(\omega) := \lim \inf_{t\to\infty} X_t(\omega)$$

and similarly $$X^*(\omega) := \lim \sup_{t\to\infty} X_t(\omega).$$

Note that both are well-defined since the martingale is assumed to have a uniform bound. Then from the assumption that X_∞ is not almost everywhere defined, we have

$$\lambda\{\omega \in \Omega \mid X_*(\omega) < X^*(\omega)\} > 0.$$

Take two rational numbers $r < s$. Let

$$B(r,s) := \{\omega \in \Omega \mid X_*(\omega) < r < s < X^*(\omega)\}.$$

Since $\{\omega \in \Omega \mid X_*(\omega) < X^*(\omega)\}$ is the countable union of all such sets $B(r,s)$, from the subadditivity of λ it follows that $\lambda(B(r^*,s^*)) > 0$ for some r^* and s^*. Then it is clear that on $B(r^*,s^*)$ the number of upcrossings U_n over the interval (r^*,s^*) increases to infinity as $n \to \infty$. In particular this implies that $\mathbb{E}(U_n) \to \infty$. However, in Lemma 12 we have seen that $\mathbb{E}(U_n)$ is bounded by $\frac{2}{s-r}\max\{|r|, K\}$. Contradiction. ◁

APPENDIX D. CONTINUITY OF THE BAYES OPERATOR

In this section we will show that the Bayes operator defined in section 3 is continuous. First of all, notice that the denominator in its definition is larger than zero by Lemma 16. So, it is easy to see that $B(\mu_i)(x_i)$ is a non-negative function on the Borel σ-algebra on Θ_i. The σ-additivity of $B(\mu_i)(x_i)$ follows from the σ-additivity of the integral and finally it is obvious that $B(\mu_i)(x_i)(\Theta_i) = 1$. So, $B(\mu_i)(x_i)$ is indeed a probability measure, and the Bayes operator thus only takes on values in $\mathbb{P}(\Theta)$. Back to our aim, its continuity, we first need to establish some technicalities.

Lemma 13. *Suppose that we have a sequence $(p_{ik})_{k=1}^{\infty}$ that converges to some p_i. Then there is a number K such that for all $k > K$, all x_i and all θ_i*

$$\|\pi_i(p_{ik}, x_i)g_i(x_i \mid p_{ik}, \theta_i) - \pi_i(p_i, x_i)g_i(x_i \mid p_i, \theta_i)\| < \varepsilon.$$

Proof. Suppose not. Then for every number n there is a number $k(n) \geq n$ and points $x_i(n)$ and $\theta_i(n)$ such that

$$\|\pi_i(p_{ik(n)}, x_i(n))g_i(x_i(n) \mid p_{ik(n)}, \theta_i(n)) - \pi_i(p_i, x_i(n))g_i(x_i(n) \mid p_i, \theta_i(n))\| \geq \varepsilon.$$

Since both X_i and Θ_i are compact we may assume w.l.o.g that the sequence $x_i(n)_{n=1}^{\infty}$ converges to a point x_i and the sequence $\theta_i(n)_{n=1}^{\infty}$ converges to a point

θ_i. However, since $k(n) \geq n$ by construction, we know that $p_{ik(n)} \to p_i$. Hence, taking limits yields

$$0 = \|\pi_i(p_i, x_i) g_i(x_i \mid p_i, \theta_i) - \pi_i(p_i, x_i) g_i(x_i \mid p_i, \theta_i)\| \geq \varepsilon$$

which is a contradiction. ◁

Lemma 14. *The function $p_i \colon \mathbb{P}(\Theta_i) \to P_i$ is continuous.*

Proof. Part (i). First we will show that the expected payoff function

$$\Pi_i \colon P_i \times \mathbb{P}(\Theta_i) \to \mathbb{R}$$

is continuous. Of course we suppose that $\mathbb{P}(\Theta_i)$ is endowed with the weak topology. Notice that his topology is metrizable by Theorem 11. Therefore it is sufficient to establish convergence of Π_i over sequences. So, take a sequence $(p_{ik}, \mu_{ik}) \to (p_i, \mu_i)$. We want to show that, given $\varepsilon > 0$, there exists a natural number K, such that for all $k \geq K$,

$$\|\Pi_i(p_{ik}, \mu_{ik}) - \Pi_i(p_i, \mu_i)\| \leq 2\varepsilon.$$

By the triangle inequality we only need to show that

$$\|\Pi_i(p_{ik}, \mu_{ik}) - \Pi_i(p_i, \mu_{ik})\| + \|\Pi_i(p_i, \mu_{ik}) - \Pi_i(p_i, \mu_i)\| \leq 2\varepsilon$$

for sufficiently large k. We will show that both terms on the left hand side of the inequality sign are smaller than or equal to ε for sufficiently large k. The first term reads

$$\begin{aligned}
& \|\Pi_i(p_{ik}, \mu_{ik}) - \Pi_i(p_i, \mu_{ik})\| \\
= & \left\| \int_{\Theta_i} \int_{X_i} \Big[\pi_i(p_{ik}, x_i) g_i(x_i, \theta_i \mid p_{ik}) - \pi_i(p_i, x_i) g_i(x_i, \theta_i \mid p_i) \Big] d\nu_i d\mu_{ik} \right\| \\
\leq & \int_{\Theta_i} \int_{X_i} \|\pi_i(p_{ik}, x_i) g_i(x_i, \theta_i \mid p_{ik}) - \pi_i(p_i, x_i) g_i(x_i, \theta_i \mid p_i)\| \, d\nu_i d\mu_{ik}.
\end{aligned}$$

Now take K as in Lemma 13. Then, since ν_i and μ_{ik} are all probability measures, for each $k \geq K$ the latter expression is smaller than or equal to

$$\int_{\Theta_i} \int_{X_i} \varepsilon \mathbb{1}_{\Theta_i \times X_i} d\nu_i d\mu_{ik} = \varepsilon.$$

Furthermore, the second term reads

$$\left\| \int_{\Theta_i} \int_{X_i} \pi_i(p_i, x_i) g_i(x_i, \theta_i \mid p_i) d\nu_i d\mu_{ik} - \int_{\Theta_i} \int_{X_i} \pi_i(p_i, x_i) g_i(x_i, \theta_i \mid p_i) d\nu_i d\mu_i \right\|.$$

Because we assume that $\mathbb{P}(\Theta_i)$ is endowed with the weak topology, it suffices to show that

$$F_p(\theta_i) := \int_{X_i} \pi_i(p_i, x_i) g_i(x_i \mid p_i, \theta_i) d\nu_i$$

is continuous in θ_i. To that end, take a sequence $\theta_{im} \to \theta_i$. Let $\varepsilon > 0$ be an arbitrary real number. Let G_p be a positive real number such that

$$\|\pi_i(p_i, x_i)\| \leq G_p \quad \text{for all } x_i \in X_i.$$

This number exists because $\pi_i(p_i, x_i)$ is continuous in x_i and X_i is compact. Now take a natural number M_p such that for all $m \geq M_p$

$$\|g_i(x_i \mid p_i, \theta_{im}) - g_i(x_i \mid p_i, \theta_i)\| \leq \frac{\varepsilon}{G_p}.$$

Then for all $m \geq M_p$

$$\begin{aligned} \|F_p(\theta_{im}) - F_p(\theta_i)\| &= \| \int_{X_i} \pi_i(p_i, x_i) \Big(g_i(x_i \mid p_i, \theta_{im}) - g_i(x_i \mid p_i, \theta_i) \Big) d\nu_i \| \\ &\leq \int_{X_i} \|\pi_i(p_i, x_i)\| \|g_i(x_i \mid p_i, \theta_{im}) - g_i(x_i \mid p_i, \theta_i)\| d\nu_i. \end{aligned}$$

Consequently, since $\nu_i(X_i) = 1$,

$$\|F_p(\theta_{im}) - F_p(\theta_i)\| \leq \int_{X_i} G_p \frac{\varepsilon}{G_p} d\nu_i = \varepsilon.$$

Part (ii). Now let $(\mu_{ik})_{k=1}^{\infty}$ be a sequence converging to μ_i in the weak topology. Then, since P_i is a compact metric space, every sequence has a converging subsequence by Theorem 12. So, we may assume without loss of generality that $p_i(\mu_{ik})$ converges to some decision p_i^*. We will now show that $p_i^* = p_i(\mu_i)$.

Since $p_i(\mu_{ik})$ is the optimal decision given the belief μ_{ik}, we know that for an arbitrary p_i in P_i it holds that

$$\Pi_i(p_i(\mu_{ik}), \mu_{ik}) \geq \Pi_i(p_i, \mu_{ik}) \quad \text{for all } k.$$

So by the continuity of Π_i we get that

$$\Pi_i(p_i^*, \mu_i) \geq \Pi_i(p_i, \mu_i),$$

and p_i^* is an optimal action given belief μ_i since p_i was arbitrarily chosen. Hence, $p_i^* = p_i(\mu_i)$ by Assumption 1. $\triangleleft$

Furthermore, notice that $g_i: X_i \times P_i \times \Theta_i \to \mathbb{R}$ is also continuous. So, the function $h: X_i \times \mathbb{P}(\Theta_i) \times \Theta_i \to \mathbb{R}$ by

$$h(x_i, \mu_i, \theta_i) := g_i(x_i, p_i(\mu_i), \theta_i)$$

is continuous as well. Now suppose we have a sequence $(x_{in}, \mu_{in})_{n=1}^{\infty}$ converging to some limit (x_i, μ_i). Define the functions f_n and f from Θ_i to $\mathbb{R}$ by

$$f_n(\theta_i) := h(x_{in}, \mu_{in}, \theta_i) \quad \text{and} \quad f(\theta_i) := h(x_i, \mu_i, \theta_i).$$

Now take an arbitrary $\varepsilon > 0$. We then have the following lemma.

Lemma 15. *There exists a natural number N in $\mathbb{N}$ such that $\|f_n - f\|_\infty < \varepsilon$ for all $n \geq N$.*

Proof. Suppose not. Then there is a subsequence $(f_k)_{k=1}^{\infty}$ of $(f_n)_{n=1}^{\infty}$ and a sequence $(\theta_{ik})_{k=1}^{\infty}$ such that

$$|f_k(\theta_{ik}) - f(\theta_{ik})| \geq \varepsilon.$$

for all $k \in \mathbb{N}$. Since Θ_i is compact we may assume that the sequence $(\theta_{ik})_{k=1}^{\infty}$ converges to some limit θ_i. Then for all $k \in \mathbb{N}$

$$\varepsilon \leq |f_k(\theta_{ik}) - f(\theta_{ik})| = |h(x_{ik}, \mu_{ik}, \theta_{ik}) - h(x_i, \mu_i, \theta_{ik})|.$$

However, since $x_{ik} \to x_i$, $\mu_{ik} \to \mu_i$ and $\theta_{ik} \to \theta_i$, the continuity of h yields

$$\varepsilon \leq |h(x_i, \mu_i, \theta_i) - h(x_i, \mu_i, \theta_i)|$$

so that we arrive at a contradiction. $\triangleleft$

The lemma is instrumental in the proof of the following

Theorem 19. *The Bayes operator is continuous.*

Proof. Suppose that $(\mu_{in}, x_{in})_{n=1}^{\infty}$ converges to (μ_i, x_i). It has to be shown that

$$B_i(\mu_{in})(x_{in}) \to B_i(\mu_i)(x_i)$$

as n goes to infinity. It is sufficient to establish (3) of Theorem 16. To this end, let F be a closed μ_i-continuous subset of Θ_i. What has to be shown is that

$$B_i(\mu_{in})(x_{in})(F) \to B_i(\mu_i)(x_i)(F).$$

By the definition of the Bayes operator,

$$B_i(\mu_i)(x_i)(F) = \frac{\int_F g_i(x_i \mid p(\mu_i), \theta_i) d\mu_i}{\int_{\Theta_i} g_i(x_i \mid p(\mu_i), \theta_i) d\mu_i}.$$

Now Lemma 16 in Appendix E guarantees that the denominator is strictly positive. So, since Θ_i itself is an instance of a closed set F whose boundary has measure zero (the boundary of Θ_i is the empty set after all), it suffices in turn to show that, given $\varepsilon > 0$,

$$\left| \int_F g_i(x_{in} \mid p(\mu_{in}), \theta_i) d\mu_{in} - \int_F g_i(x_i) \mid p(\mu_i), \theta_i) d\mu_i \right| < 2\varepsilon$$

for sufficiently large n. This is what we set out to do.

First, take N as in Lemma 15. Then for all $n \geq N$,

$$\begin{aligned} & \left| \int_F g_i(x_{in} \mid p(\mu_{in}), \theta_i) d\mu_{in} - \int_F g_i(x_i \mid p(\mu_i), \theta_i) d\mu_{in} \right| \\ \leq & \int_F |f_n(\theta_i) - f(\theta_i)| \, d\mu_{in} \leq \int_{\Theta_i} |f_n(\theta_i) - f(\theta_i)| \, d\mu_{in} \\ \leq & \int_{\Theta_i} \|f_n - f\|_\infty d\mu_{in} \leq \varepsilon \mu_{in}(\Theta_i) = \varepsilon, \end{aligned}$$

where the last inequality follows from the choice of n and N. So now we only have to show that for all $\varepsilon > 0$, there exists an $N \in \mathbb{N}$ such that for all $n \geq N$

$$\left| \int_F g_i(x_i \mid p(\mu_i), \theta_i) d\mu_{in} - \int_F g_i(x_i \mid p(\mu_i), \theta_i) d\mu_i \right| < \varepsilon.$$

In other words, we have to show the existence of a natural number N such that for every $n \geq N$, μ_{in} is an element of the set of probability measures ν_i with

$$\left| \int_F f(\theta_i) d\nu_i - \int_F f(\theta_i) d\mu_i \right| < \varepsilon.$$

This set however contains an element C of $\mathtt{C}$ with $\mu_i \in C$ by Lemma 5. Hence, such an N exists since $(\mu_{in})_{n=1}^{\infty}$ weakly converges to μ_i and $\mathtt{C}$ is a basis of the weak topology by Lemma 6. ◁

APPENDIX E. SUFFICIENTLY WIDE WORLD VIEWS

For the Bayesian learning process to be well specified, we need that there are no objectively possible events that are assigned probability zero at any time by the firm. A Bayesian learner, namely, would simply not be able to deal with such events. Formally it means that the denominator of the updating rule might become zero. In this section we will show that Assumption 2 avoids this problem. Although also several somewhat weaker conditions would guarantee that the Bayesian learning process is well defined, we prefer to work with the above condition because of its simplicity. And that it is indeed sufficient is expressed in

Lemma 16. *Let μ_i be a belief in $\mathbb{P}(\Theta_i)$ and suppose that Assumption 2 holds. Let further a decision p_i, a demand x_i, and a Borel set $A \subset \Theta_i$ with $\mu_i(A) > 0$ be given. Then*

$$\int_A g_i(\theta_i \mid p_i, x_i) d\mu_i > 0.$$

Proof. Take a decision p_i and a demand x_i. Then we know that $g_i(\theta_i \mid p_i, x_i)$ is a continuous function in the variable θ_i since we even assumed that g_i is continuous in all three variables together. Moreover, Θ_i is compact. So, there exists a real number $\varepsilon > 0$ such that $g_i(\theta_i \mid p_i, x_i) \geq \varepsilon$ for all $\theta_i \in \Theta_i$. Consequently,

$$\int_A g_i(\theta_i \mid p_i, x_i) d\mu_i \geq \int_A \varepsilon 1_{\Theta_i} d\mu_i = \varepsilon \int_A 1_{\Theta_i} d\mu_i = \varepsilon \mu_i(A)$$

which is positive since both ε and $\mu_i(A)$ are positive by assumption. ◁