TOPICS IN MICROECONOMICS: DYNAMICS AND LEARNING

MAX STINCHCOMBE

1. Introduction

There is a small number of limit theorems at the heart of theoretical studies of learning and dynamics. I want you to read and understand the major results in the theory of learning in games that are based on these limit theorems. We will therefore cover quite a bit of analysis, probability theory, and stochastic process theory. There will be a common set of required homeworks for the course, and a number of possible Detours you can take according to your interests. You should choose to take two of the Detours, and if you are interested in a different detour more closely aligned with your interests, suggest it to me and we'll arrange it.

Here is a rough outline of the course, including some (but not all of the detours):

1. Introduction.
2. Sequence Spaces: These are the crucial mathematical constructs for the limit theorems that are behind learning theory. Deterministic dynamic systems give rise to points in sequence spaces, statistical learning and stochastic process theory can be studied as probabilities on sequence spaces.
3. Metric Spaces:
   (a) Completeness, the metric completion theorem.
   (b) Constructing $\mathbb{R}$ and $\mathbb{R}^k$.

      Detour: the contraction mapping theorem; stability conditions for deterministic dynamic systems; exponential convergence to the unique ergodic distribution of a finite, communicating Markov chain; the

Date: Fall 2001.

existence and uniqueness of a value function for discounted dynamic programming.

(c) Compactness.

Detour: Berge's theorem of the maximum; continuity of value functions; upper-hemicontinuity of solution sets and equilibrium sets.

(d) Fictitious play and Cesaro (non-)convergence in $\mathbb{R}^k$.

4. Probabilities of fields and σ-fields:
   (a) Finitely additive probabilities are not enough.

   Detour: money pumps and finitely additive probabilities; countably additive extensions on compactifications, [25].

   (b) Extensions of probabilities through the metric completion theorem.

   Detour: weak and norm convergence of probabilities on metric spaces; equilibrium existence and equilibrium refinement for compact metric space games.

   Detour: convergence to Brownian motion; a.e. continuous functions of weakly convergent sequences; limit distributions based on Brownian motion functionals, [6].

   (c) The Borel-Cantelli lemmas.
   (d) The tail σ-field and the 0-1 law.
   (e) Conditional probabilities, the tail σ-field, and learnability.
   (f) The martingale convergence theorem.
5. Learning in games.
   (a) Kalai and Lehrer [14] through Blackwell and Dubins' merging of opinions theorem, Nachbar's [18] response.
   (b) Hart and Mas-Colell's [10] convergence to correlated equilibria through Blackwell's [4] approachability theorem.
   (c) Self-confirming equilibria [9] of extensive form games.
   (d) The evolution of conventions, Young [28] and KMR [16] approaches, Bergin's [2] response.
   (e) Evolutionary dynamics and strategic stability [20].

2. Sequence Spaces in Selected Examples

In order for there to be something to learn about, situations, modeled here as games, must be repeated many times. Rather than try to figure out exactly what we mean by "many times," we send the number of times to infinity and look at what this process leads to. The crucial mathematical construct is a sequence space. We will also have use of the more general notion of a product space.

2.1. **Sequence spaces.** Let S be a set of points, e.g. $S = \{H, T\}$ when we are flipping a coin, $S = \mathbb{R}^2_+$ or $S = [0, M]^2$ when we are considering quantity setting games with two players, or $S = \times_{i \in I} A_i$ when we are consider repeating a game with player set I and each $i \in I$ has action set A_i.

Definition 2.1. *An* **infinite sequence***, or simply* **sequence***, in S is a mapping from $\mathbb{N}$ to S.*

A sequence is denoted many ways, $(s_n)_{n \in \mathbb{N}}$ and $(s_n)_{n=1}^{\infty}$ being the two most frequently used, sometimes (s_n) or even s_n are used too, this last one is particularly bad, s_n is the n'th element of the sequence $(s_n)_{n \in \mathbb{N}}$. Let S^∞ be the space of all sequences in S, a point $s \in S^\infty$ is of the form

$$s = (z_1(s), z_2(s), \dots). \qquad (1)$$

For each $k \in \mathbb{N}$ and $s \in S$, $z_k(s) \in S$ is the k'th component of s. The $z_k : S^\infty \to S$ are called many things, including the **coordinate functions**, **natural projections**, **projections**.[1]

$S^n = S \times \cdots \times S$ is the n-fold Cartesian product of S; it consists of n-length sequences[2] $(u_1, \dots, u_n)$ of elements of S. From this point of view, S^∞ is an infinite dimensional Cartesian product.

[1]Some of the basics of sequence spaces are covered in [3, Ch. 1,§2].

[2]Finite sequences will be explicitly noted, otherwise you can assume sequences are infinite.

We will often have occasion to look at spaces of the form $\Theta \times S^\infty$. A point $\theta \in \Theta$ will be an initial value for a dynamic system or a parameter of some process that is to be "learned."

2.2. **Cournot dynamics.** Two firms selling a homogeneous product to a market described by a known demand function and using a known technology decide on their quantities, $s^i \in [0, M]$, $i = 1, 2$. There is an initial state $\theta_0 = (\theta_{i,0})_{i \in I} = (\theta_{1,0}, \theta_{2,0}) \in S^2$. When t is an odd period, player 1 changes $\theta_{1,t-1}$ to $\theta_{1,t} = Br_1(\theta_{2,t-1})$, when t is an even period, player 2 changes $\theta_{2,t-1}$ to $\theta_{2,t} = Br_2(\theta_{1,t-1})$. Or, if you want to combine the periods,

$$(\theta_{1,t-1}, \theta_{2,t-1}) \mapsto (Br_1(\theta_{2,t-1}), Br_2(Br_1(\theta_{2,t-1}))).$$

In either case, note that if we set $S^0 = \{h^0\}$ (some singleton set), we have specified a **dynamic system**, that is, a class of functions $f_t : \Theta \times S^{t-1} \to S$, $t \in \mathbb{N}$. When we combine periods, the f_t has a form that is independent of the period, $f_t \equiv f$, and we have a **stationary dynamic system**. Whatever dynamic system we study, for each θ_0, the result is the outcome point

$$\mathbb{O}(\theta_0) = (\theta_0, f_1(\theta_0), f_2(\theta_0, f_2(\theta_0)), \dots),$$

a point in $\Theta \times S^\infty$. When $f_t \equiv f$ is independent of t and depends only on the previous period's outcome,

$$\mathbb{O}(\theta_0) = (\theta_0, f(\theta_0), f(f(\theta_0)), f(f(f(\theta_0))), \dots).$$

Definition 2.2. *A point $\hat{s}$ is* **stable** *for the dynamic system $(f_t)_{t \in \mathbb{N}}$ if $\exists \theta_0$ such that*

$$\mathbb{O}(\theta_0) = (\theta_0, \hat{s}, \hat{s}, \dots).$$

With the best response dynamics specified above, the stable points are exactly the Nash equilibria.

2.2.1. *Convergence, stability, and local stability.* Suppose we have a way to measure the distance between points in S, e.g. $d(u, v) = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2}$ when $S = [0, M]^2$. The d-ball around u with radius ϵ is the set $B(u, \epsilon) = \{v \in S : d(u, v) < \epsilon\}$.

Homework 2.1. *A* **metric on a set** X *is a function* $d : X \times X \to \mathbb{R}_+$ *with the following three properties:*

1. $(\forall x, y \in X)[d(x, y) = d(y, x)]$,
2. $(\forall x, y \in X)[d(x, y) = 0$ *iff* $x = y]$, *and*
3. $(\forall x, y, z \in X)[d(x, y) + d(y, z) \geq d(x, z)]$.

Show that $d(u, v) = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2}$ *is a metric on the set* $S = [0, M]^2$. *Also show that* $\rho(u, v) = |u_1 - v_1| + |u_2 - v_2|$ *and* $r(u, v) = \max\{|u_1 - v_1|, |u_2 - v_2|\}$ *are metrics on* $S = [0, M]^2$. *In each case, draw* $B(u, \epsilon)$.

There are at least two useful visual images for convergence: points s_1, s_2, s_3, etc. appearing clustered more and more tightly around u; or, looking at the graph of the sequence (remember, a sequence is a function) with $\mathbb{N}$ on the horizontal axis and S on the vertical, as you go further and further to the right, the graph gets closer and closer to u. Convergence is a crucial tool for what we're doing this semester.

Definition 2.3. *A sequence* $(s_n) \in S^\infty$ **converges to** $u \in S$ **for the metric** $d(\cdot, \cdot)$ *if for all* $\epsilon > 0$, $\exists N$ *such that* $\forall n \geq N$, $d(s_n, u) < \epsilon$. *A sequence* **converges** *if it converges to some* u.

In other notations, $s \in S^\infty$ converges to $u \in S$ if

$$(\forall \epsilon > 0)(\exists K)(\forall k \geq K)[d(z_k(s), u) < \epsilon],$$

$$(\forall \epsilon > 0)(\exists K)(\forall k \geq K)[z_k(s) \in B(u, \epsilon)].$$

These can be written $\lim_k z_k(s) = u$, or $\lim_n s_n = u$, or $z_k(s) \to u$, or $s_n \to u$, or even $s \to u$.

Example 2.1. *Some convergent sequences, some divergent sequences, and some cyclical sequences that neither diverge nor converge.*

There is yet another way to look at convergence, based on cofinal sets. Given a sequence s and an $N \in \mathbb{N}$, define the cofinal set $C_N = \{s_n : n \geq N\}$, that is, the values of the sequence from the N'th onwards. $s_n \to u$ iff $(\forall \epsilon > 0)(\exists M)(\forall N \geq$

$M)[C_N \subset B(u, \epsilon)]$. This can be said "$C_N \subset B(u, \epsilon)$ for all large N" or "$C_N \subset B(u, \epsilon)$ for large N." In other words, the English phrases "for all large N" and "for large N" have the specific meaning just given.

Another verbal definition is that a sequence converges to u if and only if it gets and stays arbitrarily close to u.

Homework 2.2. *Show that $s \in S^\infty$ converges to u in the metric $d(\cdot, \cdot)$ of Homework 2.1 iff it converges to u in the metric $\rho(\cdot, \cdot)$ iff it converges to u in the metric $r(\cdot, \cdot)$.*

Convergence is what we hope for in dynamic systems, if we have it, we can concentrate on the limits rather than on the complicated dynamics. Convergence comes in two flavors, local and global.

Definition 2.4. *A point $\hat{s} \in S$ is* **asymptotically stable** *or* **locally stable** *for a dynamic system $(f_t)_{t \in \mathbb{N}}$ if it is stable and $\exists \epsilon > 0$ such that for all $\theta_0 \in B(\hat{\theta}, \epsilon)$, $\mathbb{O}(\theta_0) \to \hat{s}$.*

Example 2.2. *Draw graphs of non-linear best response functions for which there are stable points that are not locally stable.*

When the f_t's are fixed, differentiable functions, there are derivative conditions that guarantee asymptotic stability. These results are some of the basic limit theorems referred to above.

Definition 2.5. *A point $\hat{s} \in S$ is* **globally stable** *if it is stable and $\forall \theta_0$, $\mathbb{O}(\theta_0) \to \hat{s}$.*

NB: If there are many stable points, then there cannot be a globally stable point.

2.2.2. *Subsequences, cluster points, and ω-limit points.* Suppose that $\mathbb{N}'$ is an infinite subset of $\mathbb{N}$. $\mathbb{N}'$ can be written as

$$\mathbb{N}' = \{n_1, n_2, \dots\}$$

where $n_k < n_{k+1}$ for all k. Using $\mathbb{N}'$ and sequence $(s_n)_{n \in \mathbb{N}}$, we can generate another sequence, $(s_{n_k})_{k \in \mathbb{N}}$. This new sequence is called a **subsequence** of $(s_n)_{n \in \mathbb{N}}$. The trivial subsequence has $n_k = k$, the even subsequence has $n_k = 2k$, the odd has $n_k = 2k - 1$, the prime subsequence has n_k equal to the k'th prime integer, etc.

Definition 2.6. *A* **subsequence of** $s = (s_n)_{n\in\mathbb{N}}$ *is the restriction of s to an infinite* $\mathbb{N}' \subset \mathbb{N}$.

By the one-to-one, onto mapping $k \leftrightarrow n_k$ between $\mathbb{N}$ and $\mathbb{N}'$, every subsequence is a sequence in its own right. Therefore we can take subsequences of subsequences, subsequences of subsequences of subsequences, and so on.

Sometimes a subsequence of (s_n) will be denoted $(s_{n'})$, think of $n' \in \mathbb{N}'$ to see why the notation makes sense.

Definition 2.7. *u is a* **cluster point** *or* **accumulation point** *of the sequence* $(s_n)_{n\in\mathbb{N}}$ *if there is a subsequence* $(s_{n_k})_{k\in\mathbb{N}}$ *converging to u.*

s_n converges to u iff for all $\epsilon > 0$, the cofinal sets $C_N \subset B(u, \epsilon)$ for all large N. s_n clusters or accumulates at u iff for all $\epsilon > 0$, the cofinal sets $C_N \cap B(u, \epsilon) \neq \emptyset$ for all large N. Intuitively, u is a cluster point if the sequence visits arbitrarily close to u infinitely many times, and u is a limit point if the sequence does nothing else.

Example 2.3. *Some convergent sequences, some cyclical sequences that do not converge but cluster at some discrete points, a sequence that clusters "everywhere."*

Let $\text{accum}(s)$ be the set of accumulation points of an $s \in S^\infty$.

Definition 2.8. *The set of* **ω-limit points** *of the dynamic system* $(f_t)_{t\in\mathbb{N}}$ *is set*

$$\bigcup_{\theta\in\Theta} \text{accum}(\mathbb{O}(\theta)).$$

If a dynamic system cycles, it will have ω-limit points. Note that this is true even if the cycles take different amounts of time to complete.

Example 2.4. *A straight-line cobweb example of cycles, curve the lines outside of some region to get an attractor.*

The distance between a set S' and a point x is defined by $d(x, S') = \inf\{d(x, s') : s' \in S'\}$ (we will talk in detail about inf later, for now, if you haven't seen it, treat it as a min). For $S' \subset S$, $B(S', \epsilon) = \{x : d(x, S') < \epsilon\}$. If you had graduate micro from me, you've seen this kind of set.

When $\Theta = S$ and S is compact, a technical condition that we will spend a great deal of time with (later), we have

Definition 2.9. *A set $S' \subset S$ is* **invariant** *under the dynamical system $(f_t)_{t\in\mathbb{N}}$ if $\theta \in S'$ implies $\forall k$, $z_k(\mathbb{O}(\theta)) \in S'$. An invariant S' is an* **attractor** *if $\exists \epsilon > 0$ such that for all $\theta \in B(S', \epsilon)$, $\text{accum}(\mathbb{O}(\theta)) \subset S'$.*

Strange attractors are really cool, but haven't had much impact in the theory of learning in games, probably because they are so strange.

2.3. **Statistical learning.** Estimators, which are themselves random variables, are consistent if they converge to the true value of the unknown parameter. If we think of the sampling distribution around our estimates, or, if you're a Bayesian, the posterior distribution, the change from what you knew before (either nothing or a prior) to what you now know represents learning. The convergence to the true value of the parameter is probabilistic, and typically, at any point in time, we have a probability distribution with strictly positive variance. So we haven't "learned" something we're sure of, but still, it ain't bad. This is a form of learning that has been studied for a long time. We'll look at a simple example, and then make it look more complicated.

2.3.1. *A basic statistical learning example.* Suppose that θ is uniformly distributed on $\Theta = [0, 1]$, and that $X_1, X_2, \ldots$ are i.i.d. with $P(X_n = 1) = \theta$, $P(X_n = 0) = 1-\theta$. First we pick some coin, parametrized by θ, its probability of giving 1, then we start flipping that coin repeatedly. You should have learned that

$$\overline{X}_n = n^{-1}\sum_{t=1}^{n} X_t \to \theta, \quad \text{and that} \quad n^{-\frac{1}{2}}\sum_{t=1}^{N}(X_t - \theta) \xrightarrow{w} N(0, \theta(1-\theta))$$

where "$\xrightarrow{w}$" is weak convergence or weak* convergence. Weak convergence was, most likely, defined as convergence of the cdf's. This is a special case of weak convergence, which is, more generally convergence in a special metric on the set of distributions. We will investigate it in some detail below. All those caveats aside, this is the sense in which we can learn θ.

Consider the mapping $\theta \mapsto P_\theta$ where P_θ is the distribution $P_\theta(X_n = 1) = \theta$, $P_\theta(X_n = 0) = 1 - \theta$. Repeating what we can learn in a different way:

> If we know that some random $\theta \in [0, 1]$ is drawn and then we see a sequence of i.i.d. P_θ random variables, we can learn θ, equivalently, we can learn P_θ.

This learnability starts from a position of a great deal of knowledge of the structure generating the sequence of random variables. This leads to the question of what structures are learnable [13]. To get at this question, we need a detour through probabilities on S^∞ and different ways of expressing them.

2.3.2. *Probabilities on* $\{0, 1\}^\infty$. For any $\theta \in [0, 1]$, there is a probability μ_θ on $\{0, 1\}^\infty$ corresponding to the distribution over sequences given by i.i.d. P_θ draws. The process we described, pick $\theta \in [0, 1]$ at random then pick a sequence according to μ_θ gives rise to a particular, compound probability distribution, call it μ, on $\{0, 1\}^\infty$.

This is an important shift in point of view, we are now looking at distributions on the whole sequence space. This is very different than looking at simple P_θ's. We need to take a look at defining distributions on sequence spaces.

Here $S = \{0, 1\}$ is a two point space, and S^∞ is the space of all sequences of 0's and 1's. This will simplify aspects of the problem, though the approach generalizes to larger S's. A first observation is that any non-trivial space of sequences is quite large.

Definition 2.10. *A set X is* **countable** *if there is an onto function $f : \mathbb{N} \to X$. Thus, finite sets are countable, as are infinite subsets of $\mathbb{N}$. Sets that are not countable are* **uncountable***.*

Lemma 2.11. $\{0,1\}^\infty$ *is uncountable.*

Proof: Take an arbitrary $f : \mathbb{N} \to \{0,1\}^\infty$. It is sufficient to show that f is not onto. We will do this by producing a point $s \in \{0,1\}^\infty$ that is not an $f(n)$ for any n. Arrange the $f(n) \in \{0,1\}^\infty$ as follows:

n	$z_1(f(n))$	$z_2(f(n))$	$z_3(f(n))$	$z_4(f(n))$	$z_5(f(n))$	$z_6(f(n))$	$\cdots$
1	$z_1(f(1))$	$z_2(f(1))$	$z_3(f(1))$	$z_4(f(1))$	$z_5(f(1))$	$z_6(f(1))$	$\cdots$
2	$z_1(f(2))$	$z_2(f(2))$	$z_3(f(2))$	$z_4(f(2))$	$z_5(f(2))$	$z_6(f(2))$	$\cdots$
3	$z_1(f(3))$	$z_2(f(3))$	$z_3(f(3))$	$z_4(f(3))$	$z_5(f(3))$	$z_6(f(3))$	$\cdots$
4	$z_1(f(4))$	$z_2(f(4))$	$z_3(f(4))$	$z_4(f(4))$	$z_5(f(4))$	$z_6(f(4))$	$\cdots$
5	$z_1(f(5))$	$z_2(f(5))$	$z_3(f(5))$	$z_4(f(5))$	$z_5(f(5))$	$z_6(f(5))$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Now we will add 1 modulo 2, remember the rules, $0+0=0$, $0+1=1$, $1+0=1$, and $1+1=0$. Define the point s_f by $z_n(s_f) = z_n(f(n)) + 1$ modulo 2. The point s_f differs from each $f(n)$, at the very least in the n'th coordinate. ∎

Probabilities on any set X assign numbers in $[0,1]$ to subsets of X. Subsets of X are called events. The trick is to get probabilities on the right, or at least, on useful collections of events. We'll take a first step in that direction here.

2.3.3. *Probabilities on the field of cylinder sets.* Suppose we are thinking about drawing a sequence $s \in S^\infty$ at random. For any n-sequence $(u_1, \ldots, u_n)$, the set

$$\{s \in S^\infty : (z_1(s), \ldots, z_n(s)) = (u_1, \ldots, u_n)\}$$

represents the event that first n outcomes take the values $u_1, \ldots, u_n$. For $n \in \mathbb{N}$ and $H \subset S^n$, a **cylinder set** is a set of the form

$$A_H = \{s \in S^\infty : (z_1(s), \ldots, z_n(s)) \in H\}.$$

Let $\mathcal{C}^\circ$ denote the set of cylinders. It has the important property of being a **field**.

Homework 2.3. *Show that* $\mathcal{C}^\circ$ *is a field, that is,*

1. $S^\infty, \emptyset \in \mathcal{C}^\circ$,
2. *if* $A \in \mathcal{C}^\circ$, *then* $A^c = S^\infty \setminus A \in \mathcal{C}^\circ$,
3. *if* $A_1, \ldots, A_M \in \mathcal{C}^\circ$, *then* $\cap_{m=1}^M A_m \in \mathcal{C}^\circ$.

The field $\mathcal{C}^\circ$ is countable (you should see how to prove this). Further, every $s \in S^\infty$ belongs to a countable intersection of elements of $\mathcal{C}^\circ$: for each $n \in \mathbb{N}$ and $s \in S^\infty$, let $A_n(s)$ be the cylinder set

$$\{s' \in S^\infty : (z_1(s'), \ldots, z_n(s')) = (s_1, \ldots, s_n)\}.$$

Now check that $\{s\} = \cap_n A_n(s)$. Look at questions of the form "Does s belong to A?" when $A \in \mathcal{C}^\circ$. S^∞ is uncountable, but every point in S^∞ can be specified by answering only countably many such questions.

Homework 2.4. *If $\mathcal{F}^\circ$ is a field of subsets of a set X and $A_1, \ldots, A_M \in \mathcal{F}^\circ$, then $\cup_{m=1}^M A_m \in \mathcal{F}^\circ$. Further, $A_1 \setminus A_2 \in \mathcal{F}^\circ$, and $A_1 \Delta A_2 := (A_1 \setminus A_2) \cup (A_2 \setminus A_1) \in \mathcal{F}^\circ$.*

Probabilities assign numbers to elements of fields, that is, to collections of events that are a field.

Definition 2.12. *A* **finitely additive probability** *on the field $\mathcal{F}^\circ$ of subsets of a set X is a mapping $P : \mathcal{F}^\circ \to [0,1]$ satisfying the first two conditions given here, it is* **countably additive** *on the field $\mathcal{F}^\circ$ if it also satisfies the third condition:*

1. $P(X) = 1$, *and*
2. *if $A_1, \ldots, A_M$ is a* disjoint *collection of elements of $\mathcal{F}^\circ$, then $P(\cup_m A_m) = \sum_m P(A_m)$.*
3. *if $A_1 \supset A_2 \supset \cdots \supset A_n \supset A_{n+1} \supset \cdots$ and $\cap_n A_n = \emptyset$, then $\lim_n P(A_n) = 0$.*

The third condition is sometimes called "continuity from above at $\emptyset$" and can be written as "$[A_n \downarrow \emptyset] \Rightarrow [P(A_n) \downarrow 0]$." Seems mild, but it is very powerful and has a bit of a contentious past.

Back to our example, for each $\theta \in [0,1]$ and each $u = (u_1, \ldots, u_n) \in S^n$, let $A_u = \{s \in S^\infty : (z_1(s), \ldots, z_n(s)) = (u_1, \ldots, u_n)\}$, and

$$\mu_\theta(A_u) = P_\theta(u_1) \cdot P_\theta(u_2) \cdot \cdots \cdot P_\theta(u_n).$$

In the example, every S^n is finite, so that any $H \subset S^n$ is finite, and we can define

$$\mu_\theta(A_H) = \sum_{u \in H} \mu_\theta(A_u).$$

Since finite sums can be broken up in any order, each μ_θ is a finitely additive probability on $\mathcal{C}^\circ$.

Once we have some facts about compactness in place, we will show that each μ_θ is in fact countably additive, indeed, the field $\mathcal{C}^\circ$ of subsets of S^∞ is a sufficiently specialized structure that any finitely additive probability on $\mathcal{C}^\circ$ is automatically countably additive. Lest you think that this is generally true, the following finitely additive probability is not countably additive, rather, it is trying to be as close to $\frac{1}{2}$ as possible while staying strictly above $\frac{1}{2}$,

Homework 2.5. *Let $\mathcal{B}^\circ$ denote the field of subsets of $(0,1]$ consisting of the empty set and finite unions of sets of the form $(a,b]$, $0 \le a < b \le 1$. Define a $\{0,1\}$-valued function P on $\mathcal{B}^\circ$ by $P(A) = 1$ if $(\exists \epsilon > 0)[(\frac{1}{2}, \frac{1}{2}+\epsilon) \subset A]$ and $P(A) = 0$ otherwise. Show that P is a finitely additive probability that is not countably additive.*

2.3.4. *Information and nested sequences of fields.* Sometimes, you only have partial information when you make a choice. From a decision theory point of view, there is a very important result: making your choice after you get your partial information is equivalent to making up your mind ahead of time what you will do after each and every possible piece of partial information you may receive, the Bridge-Crossing Lemma. We're after something different here, the representations of information that are available through finite fields.

Suppose that $\mathcal{F}$ is a finite field of subsets of a (for now) finite set Ω with probability P defined on 2^Ω. Let $\mathfrak{P}(\mathcal{F})$ be the partition of Ω generated by $\mathcal{F}$. For any set A, function $f : \Omega \to A$ is **$\mathcal{F}$-measurable** if for all $B \in \mathfrak{P}(\mathcal{F})$, there exist an $a_B \in A$ such that $\omega, \omega' \in B$ implies $f(\omega) = f(\omega') = a_B$. Let $M(\mathcal{F}, A)$ be the set of $\mathcal{F}$-measurable functions. For a bounded $u : \Omega \times A \to \mathbb{R}$ and a probability P, consider an interesting utility maximization problem to look at is

$$V_{(u,P)}(\mathcal{F}) := \max_{f \in M(\mathcal{F},A)} \sum_{\omega} u(\omega, f(\omega))P(\omega).$$

If the field $\mathcal{G}$ is finer than the field $\mathcal{F}$, the set $M(\mathcal{G}, A)$ is larger than the set $M(\mathcal{F}, A)$. This means that $V_{(u,P)}(\mathcal{G}) \geq V_{(u,P)}(\mathcal{F})$.

It is important to understand that larger fields are more valuable because they allow more measurable functions as strategies.

Homework 2.6 (Blackwell)**.** *An expected utility maximizer is characterized by their u an their P. Their information is characterized by a field $\mathcal{F}$. Show that $\mathcal{F}'$ is a weakly finer partition that $\mathcal{F}$ if and only if for all (u, P), $V_{(u,P)}(\mathcal{F}') \geq V_{(u,P)}(\mathcal{F})$.*

Let $\mathcal{C}_t^\circ$ be the field of sets of the form

$$\{s \in S^\infty : (z_1(s), \dots, z_t(s)) \in H,\ H \subset S^t\}.$$

Homework 2.7. *Verify that*

1. *for all t, $\mathcal{C}_t^\circ$ is a field,*
2. *for all t, then $\mathcal{C}_t^\circ \subset \mathcal{C}_{t+1}^\circ$,*

A sequence of fields, $(\mathcal{F}_t)_{t\in\mathbb{N}}$, is nested if $\mathcal{F}_t \subset \mathcal{F}_{t+1}$ for all $t \in \mathbb{N}$. A nested sequence of fields is called a **filtration**.

Homework 2.8. *If $(\mathcal{F}_t)_{t\in\mathbb{N}}$ is a filtration, then $\mathcal{F}_\infty := \cup_{t\in\mathbb{N}} F_t$ is a field.*

We will see later that $\mathcal{F}_\infty$, while large, is not large enough for our purposes.

2.3.5. *Expressing μ as a convex combination of other probabilities.* Bravely assuming the integrals means something, we can define the probability μ on $\mathcal{C}^\circ$ that the process of picking θ then getting i.i.d. P_θ random variables gives rise to by

$$\mu(A) = \int_\Theta \mu_\theta(A)\, d\theta \ \text{ for any } A \in \mathcal{C}^\circ, \quad \Theta = [0, 1].$$

This expresses μ as a convex combination of the μ_θ. Each μ_θ is learnable in the sense that, if we know that some μ_θ governs the i.i.d. sequence we're seeing, then we can consistently estimate which μ_θ is at work. Having learned μ_θ means that we have information that we can use to probabilistically forecast future behavior of the system.

There are other ways to express μ, a probability on S^∞, as a convex combination of other probabilites on S^∞. For example, for any $s \in S^\infty$, define the Dirac (or point mass) probability δ_s by

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

The following is almost a repeat of something above. It helps understand why δ_s is best understood as the special kind of probability that picks s for sure.

Homework 2.9. *Show that for any $s \in S^\infty$, $\{s\} = \bigcap\{A : s \in A,\ A \in \mathcal{C}^\circ\}$.*

One view of probability is that there is no randomness in the world, that the true s has already been picked, it's just that we don't know everything that there is to be known. We can express μ that way, suppose that $\Theta = S^\infty$, and some $s \in \Theta$ is picked according to μ, and then we see draws at different times according to δ_s. It is very clear, at least intuitively, that

$$\mu(A) = \int_\Theta \delta_s(A)\, d\mu(s), \quad \Theta = S^\infty.$$

If we knew which δ_s had been picked, then we could forecast exactly what would happen in each period in the future, there would be no uncertainty. However, no finite amount of data will ever let us get close to learning δ_s. In the limit, once we've seen all of the $z_k(s)$, we'll know δ_s, but after seeing $z_k(s)$, $k = 1, \ldots, K$ for any finite K, we'll be in essentially the same ignorant position about s as we started in. Here, the limit amount of information is discontinuous.

The difference in attitude behind the two representations of μ is huge. The first one looks at something we can at least approximately learn and defines it as useful because it contains information about what is going to happen in the future. The second one looks at a perfectly functioning crystal ball that we will never have, it would be useful, but we'll never get it.

The last representation of μ was too fine, it used δ_s's. Here's a third, very coarse representation. Let μ_{high} be the probability on S^∞ that arises when we pick a θ at random and uniformly in $(\frac{1}{2}, 1]$, then see a μ_θ distribution on S^∞. In a similar fashion, let μ_{low} be the probability on S^∞ that arises when we pick a θ at random and uniformly in $(0, \frac{1}{2}]$, then see a μ_θ distribution on S^∞. It should be clear that for all cylinder sets A,

$$\mu(A) = \tfrac{1}{2}\mu_{high}(A) + \tfrac{1}{2}\mu_{low}(A).$$

It should be at least intuitively clear that both μ_{high} and μ_{low} are learnable, but that they are much coarser than the μ_θ's, which are also learnable.[3]

2.4. **The naiveté of statistical learning.** Let us not forget our game theoretic training. Suppose that player j treats player i's choice of $a_i \in A_i$ as being i.i.d. and governed by a distribution μ_i. If this is so, it seems reasonable (after all of this time studying expected utility theory) to suppose that j tries to learn μ_i and best responds to their estimates.

Now suppose that i knows that this is how j behaves. What should i do? They should consistently play that a_i that solves

$$\max\ u_i(a_i, Br_j(a_i)).$$

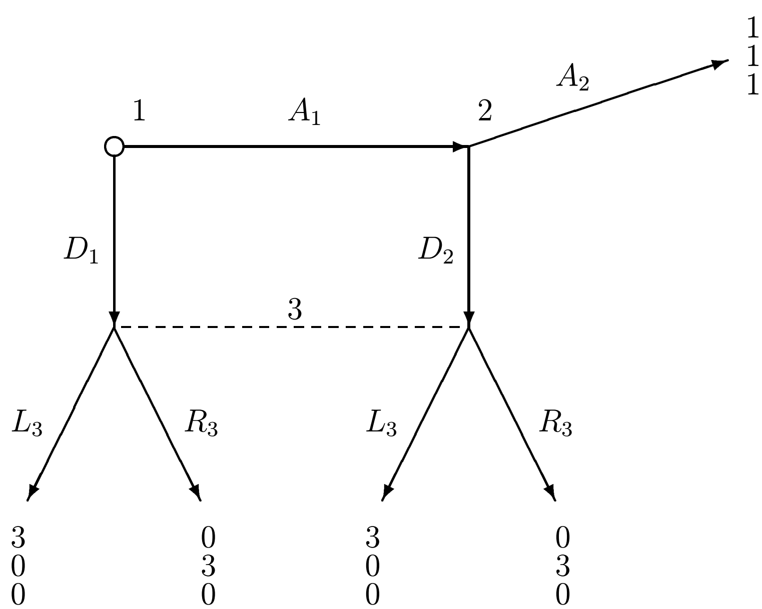
This gives them the Stackelberg payoffs to the game. In other words, i should not learn something, they should teach something.

It is this need to incorporate strategic thinking that makes the theory of learning in games so very different from statistical (and other engineering oriented) theories of learning. The tension is between mechanistic models of peoples' behavior which are, relatively speaking, easy to analyze, and models of how people think, which are, relatively speaking, very difficult to analyze. However, the tools from statistics are

[3]Think of Goldilocks and the Three Bears.

well-developed and sophisticated, we would be foolish to turn away from them just because they have not already done what we wish to do.

2.5. **Self-confirming equilibria.** Let us not forget our training in extensive form games. To analyze the equilibrium sets of an extensive form game, it is often very important to know what people will do if something they judge to be impossible, or at least very unlikely, happens. Statistical learning proceeds through the accumulation of evidence, and for reasonable people, we hope that evidence trumps theories. It is difficult to gather evidence about events that do not happen, so theories about what will happen at unreached parts of the game tree may not be so thoroughly tested by evidence. With this in mind, what can be learned? Consider the following horse game, taken from [7]:



Suppose that 1 (resp. 2) starts with the belief that 3 plays R_3 (resp. L_3) with probability greater than 2/3, and believes that 2 plays A_2 with probability close to 1. Then we expect 1 to play A_1, 2 to play A_2, and no evidence about 3's behavior will be gathered. Provided it is only evidence from observing 3's actions that goes into updating of beliefs, this means that we'll see A_1 and A_2 again in the next period, and the next, and so on. This is called a "self-confirming" equilibrium, though perhaps the non-negative "not self-denying" equilibrium would be a better term.

One way to get to the conclusion that it is only evidence from observing 3's actions that goes into updating beliefs is to assume that each player believes that the others are playing independently. If 1 thought that 2's play was correlated in some fashion with 3's play, then continuining to learn that 2's play is concentrated on A_2 could, in principle, affect 1's beliefs about 3. One story that game theorists often find plausible for this correlation involves noting that if 1 thinks that 2 is maximizing their expected utility **and 1 knows 2's payoffs**, then they learn that 2's beliefs are not in line with 1's, that someone's wrong.

So, once again, sophistication in thinking about strategic situations makes simple models of learning look too simple. But this example does a good bit more, it makes our search for Nash equilibria look a bit strange, we just gave a sensible dynamic story that has, as a stable point, even a locally stable point, strategies that are not a Nash equilibrium. The dynamic story is based on 1 and 2 having different beliefs about 3's strategy, and Nash equilibrium requires mutual best response to the same beliefs about others' strategies.

3. Metric Spaces, Completeness, and Compactness

We'll start with the most famous metric spaces, $\mathbb{R}$ and $\mathbb{R}^k$. They are complete, which is crucial. We'll also start looking at compactness in the context of these two spaces. A partial list of other metric spaces we'll look at include discrete spaces, S^∞ when S is a metric space, the set of strategies for an infinitely repeated finite game, the set of cdf's on $\mathbb{R}$,

3.1. The completeness of $\mathbb{R}$ and $\mathbb{R}^k$.

Intuitions about integers, denoted $\mathbb{N}$, are very strong, they have to do with counting things. Including 0 and the negative integers gives us $\mathbb{Z}$. The rationals, $\mathbb{Q}$, are the ratios m/n, $m, n \in \mathbb{Z}$, $n \neq 0$.

Homework 3.1. *$\mathbb{Z}$ and $\mathbb{Q}$ are countable.*

We can do all physical measurements using $\mathbb{Q}$ because they have a denseness property — if $q, q' \in \mathbb{Q}$, $q \neq q'$, then there exists a q'' half-way between q and q', i.e. $q'' = \frac{1}{2}q + \frac{1}{2}q'$ is a rational. One visual image: if we were to imagine stretching the rational numbers out one after the other, nothing of any width whatever could get through, it's an infinitely fine sieve. However, it is a sieve that, arguably, has holes in it.

One of the theoretical problems with $\mathbb{Q}$ as a model of quantities is that there are easy geometric constructions that yield lengths that do not belong to $\mathbb{Q}$ — consider the length of the diagonal of a unit square, by Pythagoras' Theorem, this length is $\sqrt{2}$.

Lemma 3.1. $\sqrt{2} \notin \mathbb{Q}$.

Proof: If $\sqrt{2} = m/n$ for some $m, n \in \mathbb{N}$, $n \neq 0$. We will derive a contradiction from this, proving the result. By cancellation, we know that at most one of the integers m and n are even. However, cross multiplying and then squaring both sides of the equality gives $2n^2 = m^2$, so it must have been m that is even. If m is even, it is of the form $2m'$ and $m^2 = 4(m')^2$ giving $2n^2 = 4(m')^2$ which is equivalent to $n^2 = 2(m')^2$, which implies that n is even, $(\Rightarrow\Leftarrow)$. ∎

If you believe that all geometric lengths must exist, i.e. you believe in some kind of deep connection between numbers that we can imagine and idealized physical measurements, this observation could upset you, and it might make you want to add some new "numbers" to $\mathbb{Q}$, at the very least to make geometry easier. The easiest way to add these new numbers is an example of a process called completing a metric space. It requires some preparation.

Definition 3.2. *A sequence q_n in $\mathbb{Q}$ is* **Cauchy** *if*

$$(\forall q > 0,\ q \in \mathbb{Q})(\exists M \in \mathbb{N})(\forall n, n' \geq M)[|x_n - x_{n'}| < q].$$

Intuitively, a Cauchy sequence is one that "settles down."

The set of all Cauchy sequences in $\mathbb{Q}$ is $\mathfrak{C}(\mathbb{Q})$.

Definition 3.3. *Two Cauchy sequences, x_n, y_n, are equivalent, written $x_n \sim_{\mathfrak{C}} y_n$, if*

$$(\forall q > 0,\ q \in \mathbb{Q})(\exists N \in \mathbb{N})(\forall n \geq N)[|x_n - y_n| < q].$$

Homework 3.2. *Check that $x_n \sim_{\mathfrak{C}} y_n$ and $y_n \sim_{\mathfrak{C}} z_n$ implies that $x_n \sim_{\mathfrak{C}} z_n$.*

Definition 3.4. *The set of* **real numbers**, $\mathbb{R}$, *is $\mathfrak{C}(\mathbb{Q})/\sim_{\mathfrak{C}}$, the set of equivalence classes of Cauchy sequences.*

For any Cauchy sequence x_n, $[x_n]$ denotes the Cauchy equivalence class. For example,

$$\sqrt{2} = [1, 1.4, 1.41, 1.414, 1.4142, 1.41421, 1.414213, \ldots].$$

The constant sequences are important, for any $q \in \mathbb{Q}$,

$$q = [q, q, q, \ldots].$$

Looking at the constant sequences shows that we have imbedded $\mathbb{Q}$ in $\mathbb{R}$.

We understood addition, subtraction, multiplication, and division for $\mathbb{Q}$, we just extend our understanding in a fashion very close to the limit construction. Specifically,

$$[x_n] + [y_n] = [x_n + y_n],\ [x_n] \cdot [y_n] = [x_n \cdot y_n],\ [x_n] - [y_n] = [x_n - y_n],$$

and, provided $[y_n] \neq [0, 0, 0, \ldots]$, $[x_n]/[y_n] = [x_n/y_n]$.

While these definitions seem correct, to be thorough we must check that if x_n and y_n are Cauchy, then the sequences $x_n + y_n$, $x_n \cdot y_n$, x_n/y_n, and $x_n - y_n$ are also Cauchy. So long as we avoid division by 0, they are.

Homework 3.3. *Show that if x_n and y_n are Cauchy sequences in $\mathbb{Q}$, then the sequences $x_n + y_n$ and $x_n \cdot y_n$ are also Cauchy.*

If a function $f : \mathbb{Q} \to \mathbb{Q}$ has the property that $f(x_n)$ is a Cauchy sequence whenever x_n is a Cauchy sequence, then $f(\cdot)$ can be extended to a function $f : \mathbb{R} \to \mathbb{R}$ by $f([x_n]) = [f(x_n)]$. For example, Homework 3.3 implies that $f(q) = P(q)$ satisfies this propert for any polynomial $P(\cdot)$. For another example, $f(q) = |q|$ satisfies this property.

We can also extend the concepts of "greater than" and "less than" from $\mathbb{Q}$ to $\mathbb{R}$. We say that a number $r = [x_n] \in \mathbb{R}$ is greater than 0 (or strictly positive) if there exists a $q \in \mathbb{Q}$, $q > 0$, such that $(\exists N \in \mathbb{N})(\forall n \geq N)[q \leq x_n]$. We say that $[x_n] > [y_n]$ if $[x_n] - [y_n]$ is strictly positive. The set of strictly real numbers is denoted $\mathbb{R}_{++}$.

We define the distance between two points in $\mathbb{Q}$ by $d(q, q') = |q - q'|$. This distance can be extended to $\mathbb{R}$ by what we just did, so that $d(r, r') = |r - r'|$.

Definition 3.5. *A metric space* (X, d) *is* **complete** *if every Cauchy sequence converges to a limit.*

Theorem 3.6. *With* $d(r, r') = |r - r'|$*, the metric space* $(\mathbb{R}, d)$ *is complete.*

This is a special case of the metric completion theorem, and we will prove it in the more abstract setting of general metric spaces.

Corollary 3.6.1. *The metric space* $(\mathbb{R}^k, \rho)$ *is complete with* ρ *being any of the following metrics:*

1. $\rho(x, y) = \sqrt{(x - y)^T (x - y)}$,
2. $\rho(x, y) = \sum_{n=1}^{k} |x_n - y_n|$, *or*
3. $\rho(x, y) = \max_n |x_n - y_n|$.

Homework 3.4. *Using Theorem* 3.6*, prove Corollary* 3.6.1*.*

3.2. **The metric completion theorem.** Let (X, d) be a metric space. (Recall that this requires that $d : X \times X \to \mathbb{R}_+$ where $d(\cdot, \cdot)$ satisfies three conditions:

1. (symmetry) $(\forall x, y \in X)[d(x, y) = d(y, x)]$,
2. (distinguishes points) $d(x, y) = 0$ if and only if $x = y$,
3. (triangle law) $(\forall x, y, z \in X)[d(x, y) + d(y, z) \geq d(x, z)]$.)

Let $\mathfrak{C}(X)$ denote the set of Cauchy sequences in X, define two Cauchy sequences, x_n and y_n, to be equivalent, $x_n \sim_{\mathfrak{C}} y_n$, if $(\forall \epsilon > 0)(\exists N \in \mathbb{N})(\forall n \geq M)[d(x_n, y_n) < \epsilon]$, and let $\widehat{X} = \mathfrak{C}(X)/\sim_{\mathfrak{C}}$. For any Cauchy sequence, x_n, $[x_n]$ denotes the Cauchy equivalence class. Each $x \in X$ is identified with $[x, x, x, \dots]$, the equivalence class of the constant sequence.

With $x = [x_n]$ and $y = [y_n]$ being two points in $\widehat{X}$, define $\widehat{d}$ on $\widehat{X} \times \widehat{X}$ by $\widehat{d}(x, y) = [d(x_n, y_n)]$. What needs to be checked is that $d(x_n, y_n)$ really is a Cauchy sequence when x_n and y_n are Cauchy. This is true, and comes directly from the triangle inequality.

Definition 3.7. *A set $S \subset X$ is* **dense** *in the metric space (X, d) if*

$$(\forall x \in X)(\forall \epsilon > 0)(\exists s \in S)[d(s, x) < \epsilon].$$

Intuitively, dense sets are "everywhere."

Theorem 3.8 (Metric completion). *$(\widehat{X}, \widehat{d})$ is a complete metric space and X is a dense subset of $\widehat{X}$.*

Proof: Fill it in. ∎

Homework 3.5. *If (X, d) is complete, then $\widehat{X} = X$, and a sequence x_n in X converges iff it is a Cauchy sequence.*

The property that Cauchy sequences converge is very important. There are a huge number of inductive constructions of an x_n that we can show is Cauchy. Knowing there is a limit in this context gives a good short-hand name for the result of the inductive construction. Some examples: the irrational numbers that help us do geometry; Brownian motion that helps us understand finance markets; value functions that help us do dynamic programming both in micro and in macro.

Going back to $\mathbb{R}$, we see that $\mathbb{Q}$ is a dense subset of the complete metric space $(\mathbb{R}, d)$ when d is defined by $d(x, y) = |x - y|$.

Definition 3.9. *A metric space (X, d) is* **separable** *if there is a countable $X' \subset X$ that is dense.*

The picture of $\mathbb{Q}$ as an infinitely fine sieve comes out as their denseness, and $\mathbb{R}$ is a separable metric space because $\mathbb{Q}$ is a countably dense subset. The holes in $\mathbb{Q}$ come out as the non-emptiness of $\mathbb{R} \setminus \mathbb{Q}$. The holes are everywhere too.

Homework 3.6. $\mathbb{R} \setminus \mathbb{Q}$ *is dense in* $\mathbb{R}$*.*

3.3. **Completeness and the infimum property.** Some subsets of $\mathbb{R}$ do not have minima, even if they are bounded, e.g. $S = (0, 1] \subset \mathbb{R}$. The concept of a greatest lower bound, also known as an infimum, fills this gap.

A set $S \subset \mathbb{R}$ is **bounded below** if there exists an $r \in \mathbb{R}$ such that for all $s \in S$, $r \leq S$. This is written as $r \leq S$. A number $\underline{s}$ is a **greatest lower bound (glb) for or infimum of** S if $\underline{s}$ is a lower bound and $s' > \underline{s}$ implies that s' is not a lower bound for S. Equivalently, that $\underline{s}$ is a glb for S if $\underline{s} \leq S$ and for all $\epsilon > 0$, there exists an $s \in S$ such that $s < \underline{s} + \epsilon$. If it exists, the glb of S is often written $\inf S$.

The supremum is the least upper bound, or lub. It is defined in the parallel fashion.

Homework 3.7. *If* $\underline{s}$ *and* $\underline{s}'$ *are glb's for* $S \subset \mathbb{R}$*, then* $\underline{s} = \underline{s}'$*. In other words, the glb, if it exists, is unique.*

Theorem 3.10. *If* $S \subset \mathbb{R}$ *is bounded below, there there exists an* $\underline{s} \in \mathbb{R}$ *such that* $\underline{s}$ *is the glb for* S*.*

Proof: Not easy, but not that hard once you see how to do it.

Let r be a lower bound for S, set $r^1 = r$, given that r^n has been defined, define r^{n+1} to be $r^n + 2^{m(n)}$ with $m(n) = \max\{m \in \mathbb{Z} : r^n + 2^m \leq S\}$ using the conventions that $\max \emptyset = -\infty$ and $2^{-\infty} = 0$. It is very easy to show that r^n is a Cauchy sequence, and that its limit is $\inf S$. ∎

An alternative development of $\mathbb{R}$ starts with $\mathbb{Q}$ and adds enough points to $\mathbb{Q}$ so that the resulting set satisfies the property that all sets bounded below have a greatest lower bound. Though more popular as an axiomatic treatment, I find the present development through the metric completion theorem to be both more intuitive and more broadly useful. It also provides an instructive parallel when it comes time to develop other models of quantities. I wouldn't overstate the advantage too much though, there are good axiomatic developments of the other models of quantities.

3.4. **Detour #1: The contraction mapping theorem.** The contraction mapping theorem will yield stability conditions for deterministic dynamic systems, conditions that reappear when you add noise, exponential convergence to the unique ergodic distributions of a finite-state, communicating Markov chain, and existence and uniqueness of value functions.

3.4.1. *The contraction mapping Theorem.* Let (X, d) be a metric space. A mapping f from X to X is a **contraction mapping** if

$$(\exists\beta \in (0,1)\,)(\forall x, y \in X)[d(f(x), f(y)) < \beta d(x, y)].$$

Lemma 3.11. *If $f : X \to X$ is a contraction mapping, then for all $x \in X$, the sequence*

$$x, f^{(1)}(x) = f(x), f^{(2)}(x) = f(f^{(1)}(x)), \ldots, f^{(n)}(x) = f(f^{(n-1)}(x)), \ldots$$

is a Cauchy sequence.

Homework 3.8. *Prove the lemma.*

A **fixed point of a mapping** $f : X \to X$ is a point x^* such that $f(x^*) = x^*$. Note that when $X = \mathbb{R}^n$, $f(x^*) = x^*$ if and only if $g(x^*) = 0$ where $g(x) = f(x) - x$. Thus, fixed point existence theorems may tell about the solutions to systems of equations.

Theorem 3.12 (Contraction mapping)**.** *If $f : X \to X$ is a contraction mapping and (X, d) is a complete metric space, then f has a unique fixed point.*

Homework 3.9. *Prove the Theorem.* [*From the previous Lemma, you know that starting at any x gives a Cauchy sequence, Cauchy sequences converge because (X, d) is complete, if $\widehat{x}$ is a limit point, then show it is a fixed point, then show that there cannot be more than one fixed point.*]

3.4.2. *Stability analysis of deterministic dynamic systems.* We'll start with stationary linear dynamic systems in $\mathbb{R}^k$. Let $\Theta = S = \mathbb{R}^k$, and let $M : \mathbb{R}^k \to \mathbb{R}^k$ be a linear mapping. We're after conditions on M equivalent to $M(\cdot)$ being a contraction mapping. This gives us information about the behavior of the dynamic system starting at x_0 and satisfying $x_{t+1} = Mx_t$. Throughout this topic, feel free to use anything and everything you know about linear algebra, i.e. do not try to go back to first principles if knowing something about determinants will save you hours of frustration.

Some preliminaries:

1. Note that $x = 0$ is a stable point for the dynamic system just specified.
2. Fix a basis for $\mathbb{R}^k$ and let M also denote the $k \times k$ matrix representation of the mapping $M(\cdot)$.
3. M is also a linear mapping from $\mathbb{C}^k$ to $\mathbb{C}^k$ where $\mathbb{C}$ is the set of complex numbers.
4. The Fundamental Theorem of Algebra says that every n-degree polynomial has n roots in $\mathbb{C}$ if we count multiplicities.
5. An upper triangular matrix T is one with the property that $T_{i,j} = 0$ if $i > j$, i.e. every entry below the diagonal is equal to 0.

Lemma 3.13 (Upper Triangular). *Show that there exist an invertible matrix B such that $M = B^{-1}TB$ where T is an upper triangular matrix.*

Homework 3.10. *Prove the Upper Triangular Lemma.*

The entries in T may be complex. In particular,

Homework 3.11. *The diagonal entries, $T_{i,i}$, in T are the eigenvalues of M.*

Viewing M as a mapping from $\mathbb{R}^k$ to $\mathbb{R}^k$, and defining the norm of a vector x by $\|x\| = \sqrt{x^T x}$, define the norm of M as

$$\|M\| = \sup\{\|Mx\| : \|x\| = 1\}.$$

Homework 3.12. *M is a contraction mapping iff $\|M\| < 1$ iff form some $n \in \mathbb{N}$, $\|M^n\| < 1$.*

If M is a contraction mapping, then the dynamic system with $x_{t+1} = Mx_t$ is globally asymptotically stable.

Homework 3.13 (Probably difficult). *M is a contraction mapping iff $\max_i |T_{i,i}| < 1$. [It might be easier to prove this if you use the Jordan canonical form rather than the upper triangular form, if you go that route, carefully state and give a citation to the theorem giving you the canonical form.]*

Homework 3.14. *Using the previous problem, find conditions on α and β such that $M = \left[\begin{smallmatrix} 0 & \alpha \\ \beta & 0 \end{smallmatrix}\right]$, is a contraction mapping. Give the intuition. [By the way, if $\alpha > 0$ and $\beta < 0$ or the reverse, the eigenvalues are imaginary.] Draw representative dynamic paths in a neighborhood of the origin for the cases of contraction mappings having*

1. *$\alpha, \beta > 0$,*
2. *$\alpha, \beta < 0$,*
3. *$\alpha > 0$, $\beta < 0$, and*
4. *$\alpha < 0$, $\beta > 0$.*

Stable points can fail to be locally stable in a number of ways. We've seen an example where nothing no starting point near the stable point converged to it. Here's another possibility.

Homework 3.15. *Draw representative dynamic paths in a neighborhood of the origin when M is the matrix $M = \left[\begin{smallmatrix} \alpha & 0 \\ 0 & \beta \end{smallmatrix}\right]$, $\alpha > 1$, $0 < \beta < 1$.*

Now let's suppose that instead of being linear, the transformation is affine, i.e. $A(x) = a + Mx$ for some $a \in \mathbb{R}^k$ and some $k \times k$ invertible matrix M.

Homework 3.16. *Show that the dynamic system with $x_{t+1} = A(x_t)$ has a unique stable point, x^*.*

Shifting the origin to x^* means treating any vector x as being the vector $x - x^*$. The next result shows that if we shift the origin to x^* and analyze the stability properties of M in the new, shifted world, we are actually analyzing the stability properties of $A(\cdot)$.

Homework 3.17 (Easy). *Show that for any x_t, $A(x_t) - x^* = Mv_t$ where $v_t = x_t - x^*$.*

Homework 3.18. *Suppose in a Cournot game, the best responses are*

$$Br_1(q_2) = \max\{0, a - bq_2\}, \text{ and } Br_2(q_1) = \max\{0, c - dq_1\},\ \ a, b, c, d > 0.$$

Analyze the stability of the dynamic system on $\mathbb{R}^2_+$

$$\begin{bmatrix} q_{1,t+1} \\ q_{2,t+1} \end{bmatrix} = \begin{bmatrix} Br_1(q_{2,t}) \\ Br_2(q_{1,t}) \end{bmatrix}.$$

Now consider dynamic systems with L lags:

$$x_t = a + \sum_{\ell=1}^{L} \beta_\ell x_{t-\ell}.$$

Homework 3.19. *For any t, let X_t be the transpose of the vector $[x_t, x_{t-1}, \ldots, x_{t-L+1}]$. Express the dynamic system just given in $L \times L$ matrix form using X_t and X_{t-1}. Give conditions on the β_ℓ's guaranteeing global asymptotic stability.*

If $(\varepsilon_t)_{t\in\mathbb{N}}$ is a sequence of i.i.d. mean 0, finite variance random variables, the stochastic dynamic system

$$x_t = a + \sum_{\ell=1}^{L} \beta_\ell x_{t-\ell} + \varepsilon_t$$

provides a model with a great deal of interesting dynamic behavior. Having the eigenvalues inside the unit circle (in the complex plane) gives (one of the many things that is called) stationary behavior. Basically, noise from the distant past keeps being contracted out of existence, but noise from the more recent past is always there. A special well-studied has an eigenvalue directly on the unit circle,

$$x_t = x_{t-1} + \varepsilon_t.$$

This is called a random walk, you get the classical random walk by starting with $x_0 = 0$ and having $\varepsilon_t = \pm 1$ with probability half apiece.

All of this linear analysis can be transplanted to non-linear systems by taking derivatives. Suppose that $f : \mathbb{R}^k \to \mathbb{R}^k$ is a twice continuously differentiable function and that $f(x^*) = x^*$ so that x^* is a stable point of the dynamic system $x_{t+1} = f(x_t)$. Giving a careful proof of the following takes a bit of doing, and may even require all the differentiability assumed. However, the idea is really primitive, we just pretend that x^* is the origin, then replace the function f by its Taylor expansion, ignore all but the first, linear terms, and show that the approximation errors don't mess anything up, even when accumulated over time.

Lemma 3.14. *If $D_x f(x^*)$ is invertible and a contraction mapping, then x^* is locally stable.*

3.4.3. *Stationary, ergodic Markov chains with finite state spaces.* We've already defined probabilities on the cylinder sets of $\{0, 1\}^\infty$, replacing $\{0, 1\}$ by any finite S doesn't change that construction in any significant way. We are now going to look at probabilities on S^∞, S finite, that are not independent.

Let P_0 be an arbitrary probability on S. For $i, j \in S$, let $P_{i,j} \geq 0$ satisfy $(\forall i \in S)[\sum_j P_{i,j} = 1]$. From these ingredients, we are going to define a probability on $S \times S^\infty$.

For any $n+1$-sequence $(u_0, u_1, \ldots, u_n)$ in $S \times S^n$, the set

$$\{(u_0, s) : s \in S^\infty,\ (z_1(s), \ldots, z_n(s)) = (u_1, \ldots, u_n)\}$$

has probability

$$P_0(u_0) \cdot P_{u_0,u_1} \cdot P_{u_1,u_2} \cdot \cdots \cdot P_{u_{n-1},u_n}.$$

Since $S \times S^n$ is finite, this gives a probability on the cylinder sets, $\mathcal{C}^\circ$. Such a probability is called a stationary Markov process.

Suppose that we draw $(s_0, s) \in S \times S^\infty$ according to such a probability. Let $X_t(s)$ be the measurable function (a.k.a. random variable) $z_t(s)$, $t = 0, 1, \ldots$. The **Markov property** is that

$$(\forall t)[P(X_{t+1} = j | X_0 = i_0, \ldots, X_{t-1} = i_{t-1}, X_t = i) = P(X_{t+1} = j | X_t = i) = P_{i,j}.$$

In words, in the history of the random variables, $X_0 = i_0, \ldots, X_{t-1} = i_{t-1}, X_t = i$, only the last period, $X_t = i$, contains any probabilistic information about X_{t+1}.

It seems that Markov chains must have small memories, after all, the distribution of X_{t+1} depends only on the state at time t. This can be "fixed" by expanding the state space, e.g. replace S with $S \times S$ and the last two realizations of the original X_t can influence what happens next.

The matrix P is called the **one-step transition matrix**. This name comes from the following observation: if π^T is the (row) vector of probabilities describing the distribution of X_t, then $\pi^T P$ is the (row) vector describing the distribution of X_{t+1}.

For $i, j \in S$, let $P^{(n)}_{i,j} = P(X_{t+n} = j | X_t = i)$. The matrix $P^{(n)}$ is called the **n-step transition matrix**. One of the basic rules for stationary Markov chains is called the Chapman-Kolmogorov equation,

$$(\forall 1 < m < n)[P^{(n)}_{i,j} = \sum_{k \in S} P^{(m)}_{ik} \cdot P^{(n-m)}_{kj}].$$

Homework 3.20. *Verify the Chapman-Kolmogorov equation.*

This means that if π^T is the (row) vector of probabilities describing the distribution of X_t, then $\pi^T P^{(n)}$ is the (row) vector describing the distribution of X_{t+n}.

Homework 3.21. *The matrix $P^{(n)}$ is really the matrix P multiplied by itself n times.*

Let $\Delta(S)$ denote the set of probabilities on S. $\pi \in \Delta(S)$ is an **ergodic distribution** if $\pi^T P = \pi^T$.

Homework 3.22. *Solve for the set of ergodic distributions for each of the following P where $\alpha, \beta \in (0,1)$:*

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \begin{bmatrix} \alpha & (1-\alpha) \\ (1-\alpha) & \alpha \end{bmatrix} \quad \begin{bmatrix} \alpha & (1-\alpha) \\ (1-\beta) & \beta \end{bmatrix} \quad \begin{bmatrix} 1 & 0 \\ (1-\beta) & \beta \end{bmatrix}$$

Theorem 3.15. *If S is finite and there exists an N such that for all $n \geq N$, $P^{(n)} \gg 0$, then the mapping $\pi^T \mapsto \pi^T P$ from $\Delta(S)$ to $\Delta(S)$ is a contraction mapping.*

Proof: For each $j \in S$, let $m_j = \min_i P^N_{i,j}$. Because $P^N \gg 0$, we know that for all j, $m_j > 0$. Define $m = \sum_j m_j$. We will show that for $p, q \in \Delta(S)$, $\|pP^N - qP^N\|_1 \leq (1-m)\|p-q\|_1$.

$$\begin{aligned}
\|pP^N - qP^N\|_1 &= \sum_{j\in S} \left| \sum_{i \in S} (p_i - q_i) P^N_{i,j} \right| \\
&= \sum_{j\in S} \left| \sum_{i \in S} (p_i - q_i)(P^N_{i,j} - m_j) + \sum_{i\in S} (p_i - q_i) m_j \right| \\
&\leq \sum_{j\in S} \sum_{i\in S} |p_i - q_i| (P^N_{i,j} - m_j) + \sum_{j\in S} m_j \left| \sum_{i\in S} (p_i - q_i) \right| \\
&= \sum_{i\in S} |p_i - q_i| \sum_{j\in S} (P^N_{i,j} - m_j) + 0 \\
&= (1-m)\|p-q\|_1,
\end{aligned}$$

where the next-to-last equality follows from the observation that $p, q \in \Delta(S)$, and the last equality follows from the observation that for all $i \in S$, $\sum_{j\in S} P^N_{i,j} = 1$, and $\sum_{j \in S} m_j = m$. This shows that P^N is a contraction mapping. Since P is a linear mapping, we're done (that's a separate step, taken above for linear maps from $\mathbb{R}^k$ to $\mathbb{R}^k$, check that it works from Δ to Δ). ∎

Homework 3.23. *Verify that this proof works so long as $\sum_j m_j > 0$, a looser condition than the one given. This condition applies, for example, to the matrix*

$$\begin{bmatrix} 1 & 0 \\ (1-\beta) & \beta \end{bmatrix},$$

where $m_1 = 1 - \beta$, $m_2 = 0$, $m = 1 - \beta$, and the contraction factor is $1 - m = \beta$.

Assuming that $\Delta(S)$ is complete (it is, we just haven't proven it yet), we now have sufficient conditions for the existence of a unique ergodic distribution.

Homework 3.24. *Under the conditions of Theorem* 3.15*, show that the matrix P^n converges and characterize the limit.*

3.4.4. *The existence and uniqueness of value functions.* A maximizer faces a sequence of interlinked decision at times $t \in \mathbb{N}$. At each t, they learn the state, s, in a state space S. Since we don't yet have the mathematics to handle integrating over larger S's, we're going to assume that S is countable. For each $s \in S$, the maximizing person has available actions $A(s)$. The choice of $a \in A(s)$ when the state is s gives utility $u(a, s)$. When the choice is made at $t \in \mathbb{N}$, it leads to a random state, X_{t+1}, at time $t+1$, according to a transition probability $P_{i,j}(a)$, at which point the whole process starts again. If the sequence $(a_t, s_t)_{t \in \mathbb{N}}$ is the outcome, the utility is $\sum_t \beta^t u(a_t, s_t)$ for some $0 < \beta < 1$. Assume that there exists a $B \in \mathbb{R}_{++}$ such that $\sup_{(a_t, s_t)_{t\in\mathbb{N}},\ a_t \in A(s_t)} |\sum_t \beta^t u(a_t, s_t)| < B$. This last happens if $u(a, s)$ is bounded, or if its maximal rate of growth is smaller than β.

One of the methods for solving infinite horizon, discounted dynamic programming problems just described is called the method of succesive approximation: one pretends that the problem has only one decision period left, and that if one ends up in state s after this last decision, one will receive $\beta V_0(s)$, often $V_0(s) \equiv 0$. Define

$$V_1(s) = \max_{a \in A(s)} u(a, s) + \beta \sum_{j \in S} V_0(j) P_{s,j}(a).$$

For this to make sense, we must assume that the maximization problem has a solution, which we do. (There are sensible looking conditions guaranteeing this, the simplest is the finiteness of $A(s)$.) More generally, once V_t has been defined, define

$$V_{t+1}(s) = \max_{a \in A(s)} u(a, s) + \beta \sum_{j \in S} V_t(j) P_{s,j}(a).$$

Again, we are assuming that for any $V_t(\cdot)$, the maximization problem just specified has a solution.

We've just given a mapping from possible value functions to other possible value functions. The point is that it's a contraction mapping.

The space $X_B = [-B, +B]^S$ is the space of all functions from S to the interval $[-B, +B]$. For $v, v' \in X$, define

$$\rho(v, v') = \sup_{s \in S} |v_s - v'_s|.$$

Homework 3.25. *ρ is a metric on X_B and the metric space (X_B, ρ) is complete.*

Define the mapping $f : X_B \to X_B$ by defining the s'th component of $f(v)$, that is, $f(v)_s$, by

$$f(v)_s = \max_{a \in A(s)} u(a, s) + \beta \sum_{j \in S} v_j P_{s,j}(a).$$

Homework 3.26. *The function f just described is a contraction mapping.*

Let v^* denote the unique fixed point of f. Let $a^*(s)$ belong to the solution set to the problem

$$\max_{a \in A(s)} u(a,s) + \beta \sum_{j \in S} v_j^* P_{s,j}(a).$$

Homework 3.27. *Using the policy $a^*(\cdot)$ at all points in time gives the expected payoff $v^*(s)$ if started from state s at time* 1.

Define $\widehat{v}(s)$ to be the supremum of the expected payoffs achievable starting at s, the supremum being taken over all possible feasible policies, $\alpha = (a_t(\cdot,\cdot))_{t \in \mathbb{N}}$,

$$\widehat{v}(s) = \sup_{(a_t(\cdot,\cdot))_{t \in \mathbb{N}}} E\left(\sum_t \beta^t u(a_t, s_t) | s_1 = s \right).$$

Homework 3.28. *For all s, $v^*(s) = \widehat{v}(s)$.*

Combining the last two problems, once you've found the value function, you are one step away from finding an optimal policy, further, that optimal policy is stationary.

3.5. **Closed sets, compact sets, and accumulation points.** We've already seen that accumulation points are a way to talk about the long term behavior of dynamic systems and learning problems. Fix a metric space (X, d), for now, you'll not go wrong in thinking of $\mathbb{R}$ or $\mathbb{R}^k$ as the metric space, but most of the proofs given here will not use any of the special structure available in $\mathbb{R}$ and $\mathbb{R}^k$.

Definition 3.16. *A set $F \subset X$ is* **closed** *if, for all sequences (s_n) in F, $\text{accum}(s_n) \subset F$.*

Thus, the closed sets are the ones that contain any accumulation points of a sequence in that set. Now, it is possible that there are sequences s_n in F with the property that $\text{accum}(s_n) = \emptyset$, and for any such (s_n), the conclusion that $\text{accum}(s_n) \subset F$ is trivial.

Example 3.1. *$F = [0, \infty) \subset \mathbb{R}$ is closed, as is $F' = \mathbb{R}^2_+ \subset \mathbb{R}^2$. The sequence $s_n = n$ is a sequence in F with no accumulation points, the sequence $s_n = (n, n)$ is a sequence in F' with no accumulation points.*

Definition 3.17. *A set $K \subset X$ is* **compact** *if, for all sequences (s_n) in K, $\text{accum}(s_n) \neq \emptyset$ and $\text{accum}(s_n) \subset K$.*

Thus, compact sets are the closed one with the property that every sequence in the set must accumulate somewhere in the set. There is a relation between compactness and properties we've seen before.

Lemma 3.18. *If X is compact, then (X, d) is a complete, separable metric space.*

Proof: Since X is compact, any Cauchy sequence (s_n) in X must have an accumulation point, call it x. Therefore some subsequence $s_{n'} \to x$. Since s_n is Cauchy, it must also converge to x (yes, there is a step missing there, a step you complete by using the triangle property of metrics). The separability comes from the following result:

> For any $\epsilon > 0$, there is a finite $X^\epsilon \subset X$ such that $(\forall x \in X)(\exists x' \in X^\epsilon)[d(x, x') < \epsilon]$.

To see why separability flows from this result, observe that the countable set $X' = \cup_n X^{1/n}$ is dense. To prove this result, pick your $\epsilon > 0$. Start an inductive procedure by picking an arbitrary $x_1 \in X$. If x_1 through x_n have been picked, then pick an arbitrary x_{n+1} from $X \backslash \cup_{i=1}^n B(x_i, \epsilon)$. If this set is empty, then set $X^\epsilon = \{x_1, \ldots, x_n\}$, otherwise continue. If we can show that this procedure must terminate, then we've produced the requisite finite X^ϵ. Suppose it does not terminate. Then it gives a sequence (x_n) with the property that $d(x_n, x_m) \geq \epsilon$ for all $n \neq m$. Since X is compact, (x_n) must have an accumulation point, call it x. For some subsequence, $d(x_{n'}, x) \to 0$, but this violates the observation that $d(x_{n'}, x_{m'}) \geq \epsilon$ for any $n' \neq m'$. ∎

The sets X^ϵ in the result above are called ϵ-nets.

To repeat, compact sets are closed and have the additional property that any sequence in them has accumulation points. You have seen many compact sets in micro and game theory.

Definition 3.19. *A subset B of $\mathbb{R}^k$ is* **bounded** *if $(\exists R \in \mathbb{R})(\forall x \in B)[x^T x \leq R]$.*

Theorem 3.20. *$K \subset \mathbb{R}^k$ is compact iff it is closed and bounded.*

This is a famous theorem, the proof only looks easy in retrospect.

Proof: Fill it in. ∎

Definition 3.21. *Let (X, d) and (Y, ρ) be two metric spaces. A function $f : X \to Y$ is* **continuous at** *x if $x^n \to x$ implies $f(x^n) \to f(x)$. A function $f : X \to Y$ is* **continuous** *if it is continuous at every $x \in X$.*

Theorem 3.22. *If $f : K \to \mathbb{R}$ is continuous and K is compact (and non-empty), then $(\exists x \in K)[f(x) = \sup\{f(y) : y \in K\}]$.*

It should be clear to you, at least by the end of the proof, that we could substitute "inf" for "sup" in the above. Note that one implication is that the function f must be bounded.

Proof: Fill it in. ∎

This theorem is the reason that demand correspondences are non-empty when preferences are continuous and $(p, w) \gg (0, 0)$.

Okay, enough of the real analysis, time to go back to probability, we'll come back to real analysis as we need it. For those of you who are interested, the next detour uses real analysis to get at the properties of some of the basic theoretical constructs in economics.

3.6. **Detour #2: Berge's Theorem of the Maximum and Upper Hemicontinuity.**

For each x in a set X, there is a set $\Phi(x)$ of choices available to a maximizer, $\Phi(x) \subset Y$. The utility function of the maximizer, $f(x, y)$, depends on both arguments. One object of interest is the value function,

$$v(x) = \sup_{y \in \Phi(x)} f(x, y).$$

Provided each $f(x, \cdot)$ is continuous and each $\Phi(x)$ is compact, this can be replaced by

$$v(x) = \max_{y \in \Phi(x)} f(x, y),$$

and the set of maximizers is non-empty,

$$\Psi(x) := \{y^* \in \Phi(x) : (\forall y' \in \Phi(x))[f(x, y^*) \geq f(x, y')]\}.$$

There is no hope that $v(\cdot)$ or $\Psi(\cdot)$ are well-behaved if $f(\cdot, \cdot)$ or $\Phi(\cdot)$ is arbitrary. A quite general set of sufficient conditions for "well-behavedness" is that $f(\cdot, \cdot)$ is jointly continuous and that $\Phi(\cdot)$ is continuous. We need to define these two terms.

Let (X, d) and (Y, ρ) be two metric spaces.

Definition 3.23. $f : X \times Y \to \mathbb{R}$ *is* **jointly continuous at** (x, y) *if* $\forall \epsilon > 0$ *there is a* $\delta > 0$ *such that for all* (x', y') *with* $d(x', x) < \delta$ *and* $\rho(y', y) < \delta$, $|f(x', y') - f(x, y)| < \epsilon$. f *is* **jointly continuous** *it is jointly continuous at all* (x, y).

Homework 3.29. *Give a function* $f : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ *such that for all* x, $f(x, \cdot)$ *is continuous and for all* y, $f(\cdot, y)$ *is continuous, but* f *is not jointly continuous.*

A mapping from points to sets is called a correspondence. To guarantee that the set of maximizers is non-empty, we are going to assume that the correspondence Φ always takes on compact values, that is, for all x, $\Phi(x)$ is a non-empty, compact subset of Y. Let $\mathfrak{K}_Y$ denote the set of **non-empty** compact subsets of Y. Correspondences can be seen as functions, in this case, $\Phi : X \to \mathfrak{K}_Y$. To talk about the continuity of $\Phi(\cdot)$ we'll use a metric on $\mathfrak{K}_Y$.

For $A, B \in \mathfrak{K}_Y$, define $c(A, B) = \inf\{\epsilon > 0 : A \subset B^\epsilon\}$ where $B^\epsilon = \{y \in Y : \inf_{b\in B} d(y, b) < \epsilon\}$. The **Hausdorff distance** between compact sets is defined by

$$d_H(A, B) = \max\{c(A, B), c(B, A)\}.$$

Homework 3.30. d_H *is a metric on* $\mathfrak{K}_Y$.

The continuity of Φ comes in three flavors, upper, lower, and full.

Definition 3.24. *A correspondence* $\Phi : X \to \mathfrak{K}_Y$ *is*

1. **upper hemicontinuous (uhc) at** x *if for all* $\epsilon > 0$ *there exists a* $\delta > 0$ *such that forall* x' *with* $d(x, x') < \delta$, $c(\Phi(x'), \Phi(x)) < \epsilon$, *is*
2. **lower hemicontinuous (lhc) at** x *if for all* $\epsilon > 0$ *there exists a* $\delta > 0$ *such that forall* x' *with* $d(x, x') < \delta$, $c(\Phi(x), \Phi(x')) < \epsilon$, *and is*
3. **continuous at** x *if it is both uhc and lhc at* x, *i.e. if* $\Phi : X \to \mathfrak{K}_Y$ *is a continuous function.*

Φ *is uhc (resp. lhc, resp. continuous) if it is uhc (resp. lhc, resp. continuous) at every* x.

Intuitively, uhc correspondences can explode at a point, $\Phi(x)$ can be much larger than the $\Phi(x')$ for $d(x', x)$ very small. In a similar way, lhc correspondences can implode at a point, but continuous correspondences can do neither.

Homework 3.31. *The Walrasian budget correspondence is continuous on* $\mathbb{R}^{L+1}_{++}$ *but not on* $\mathbb{R}^{L+1}_{+}$.

Just like functions, correspondences can be identified with their graphs.

Definition 3.25. *The* **graph of a correspondence** Φ *is the set* $gr\,\Phi = \{(x, y) : y \in \Phi(x)\}$.

By definition, a sequence (x^n, y^n) in $X \times Y$ converges to (x, y) iff $d(x^n, x) \to 0$ and $\rho(y^n, y) \to 0$.

Theorem 3.26 (Closed graph)**.** *If (Y, d) is compact, then the correspondence Φ is uhc iff $gr\,\Phi$ is a closed subset of $X \times Y$.*

Homework 3.32. *Prove the closed graph theorem.*

Homework 3.33. *Let $X = \mathbb{R}_+$ and let Y be the non-compact metric space $\mathbb{R}$. Let $\Phi(x) = \{1/x\}$ if $x > 0$ and $\Phi(0) = \{0\}$. Show that $gr\,\Phi$ is closed but that Φ is not uhc.*

The following result can be generalized in a number of ways, see [12] if you're interested.

Theorem 3.27 (Berge)**.** *If $f : X \times Y \to \mathbb{R}$ is jointly continuous and $\Phi : X \to \mathfrak{K}_Y$ is continuous, then the function*

$$v(x) = \max\nolimits_{y \in \Phi(x)} f(x, y)$$

is continuous, for all $x \in X$, the set $\Psi(x)$ defined by

$$\Psi(x) = \{y^* \in \Phi(x) : (\forall y' \in \Phi(x))[f(x, y^*) \geq f(x, y')]\}$$

is non-empty and compact, and the correspondence Ψ is upper hemicontinuous.

Homework 3.34. *Prove Berge's theorem.*

Homework 3.35. *Set $X = \mathbb{R}^{L+1}_{++}$ with tyically element (p, w), and $Y = \mathbb{R}^L_+$.*

1. *For a continuous utility function $u : \mathbb{R}^L_+ \to \mathbb{R}$, the indirect utility function $v(p, w)$ is continuous and the demand correspondence, $x(p, w)$, is upper hemicontinuous,*
2. *If the demand correspondence is single-valued, then its graph is the graph of a continuous function.*
3. *There are conditions under which these last results remain true even when u depends non-trivially on prices and wealth.*

Homework 3.36. *The profit function of a neo-classical firm may not be continuous. Explain which parts of the assumptions of Berge's theorem are violated and which are not in such cases.*

Homework 3.37 (Upper hemicontinuity of the Nash correspondence)**.** *Let $\Gamma(u)$ be the normal form game with finite strategy sets S_i for each i in the finite set I, and utilities $u \in \mathbb{R}^S$, $S = \times_i S_i$. Let $Eq(u) \subset \times_i \Delta_i$, $\Delta_i := \Delta(S_i)$, be the set of Nash equilibria for $\Gamma(u)$. Verify that for all u, the best response correspondence satisfies the conditions of Kakutani's fixed point Theorem so that $Eq(u)$ is non-empty. Show that $Eq(u)$ is compact, and that the correspondence $Eq(\cdot)$ is uhc. [Remember, closed subsets of compact sets are necessarily compact.]*

Remember the game theory notation, a game is given by $(T_i, u_i)_{i \in I}$ where T_i is player i's set of pure strategies and u_i is i's utility.

Homework 3.38 (Existence and upper hemicontinuity of Perfect equilibria)**.** *As in the problem just given, let $\Gamma(u)$ be a normal form game. For each $i \in I$, let $R_i = \{\eta_i \in \mathbb{R}^{S_i}_{++} : \sum_{s_i \in S_i} \eta_i(s_i) < 1$. For each $i \in I$ and $\eta_i \in R_i$, let*

$$\Delta_i(\eta_i) = \{\sigma_i \in \Delta_i : \sigma_i \geq \eta_i\}.$$

1. *For each* $\eta = (\eta_i)_{i\in I} \in \times_i R_i$, *the game* $(\Delta_i(\eta_i), u_i)_{i\in I}$ *has an equilibrium. Let* $Eq(u,\eta)$ *denote the set of equilibria. Show that* $Eq(u,\eta)$ *is a closed, non-empty set.* [*Proving this involves checking that the best response correspondences are non-empty valued, compact valued, convex valued, and upper hemicontinuous, then applying Kakutani's theorem, which I do not expect you to prove.*]
2. *Show that the intersection of an arbitrary collection of closed sets in a metric space* (X,d) *is closed. The closure of a set* E, $\text{cl}\, E$, *in a metric space* (X,d) *is defined as the intersection of all closed sets containing* E. *This means that* $\text{cl}\, E$ *is the smallest closed set containing* E. *Show that* $x \in \text{cl}\, E$ *iff there is a sequence* x_n *in* E *such that* $d(x_n, x) \to 0$.
3. *A set* K *in a metric space* (X,d) *is compact iff every collection of closed subsets of* K *has the* **finite intersection property**: *if* $\{F_\alpha : \alpha \in A\}$ *is a collection of closed subsets of* K *and* $\cap_\alpha F_\alpha = \emptyset$, *then* $\cap_{n=1}^N F_{\alpha_n} = \emptyset$ *for some finite set* $\{\alpha_1, \ldots, \alpha_N\} \subset A$.
4. *For* $\epsilon > 0$, *let* $E^\epsilon = \text{cl}\,\{Eq(u,\eta) : (\forall i \in I)[\sum_{s_i} \eta_i(s_i) < \epsilon]\}$. *The set of* **perfect equilibria** *for* $\Gamma(u)$ *can be defined as*

   $$Per(u) = \bigcap \{E^\epsilon : \epsilon > 0\}.$$

   Verify that $\sigma \in Per(u)$ *iff there is a sequence* $\eta_n \in \times_i R_i$, $\eta_n \to 0$, *and a sequence* $\sigma_n \in Eq(u, \eta_n)$ *such that* $\sigma_n \to \sigma$.
5. *Using the compactness of* Δ *and the previous parts of this problem, show that* $Per(u)$ *is a non-empty, closed (hence compact) subset of* Δ.
6. *Show that the correspondence* $Per(\cdot)$ *is upper hemicontinuous.*

The finite intersection property of the previous problem is a very useful way to talk about compactness. Let S be a finite set, and $\mathcal{C}^\circ$ the field of cylinder subsets of S^∞. Arguments using the finite intersection property show that every finitely additive probability on $\mathcal{C}^\circ$ is countably additive. This means, *inter alia*, that the spaces $(\{0,1\}^\infty, \mathcal{C}^\circ)$ and $((0,1], \mathcal{B}^\circ)$ are quite different (in a problem above, you showed that there are finitely additive probabilities on $\mathcal{B}^\circ$ that fail to be countably additive).

Homework 3.39 (Billingsley's Theorem 2.3)**.** *Give the finite set* S *the metric* $d(x,y) = 1$ *if* $x \neq y$ *and* $d(x,y) = 0$ *if* $x = y$. *Give the sequence space the metric* $\rho(s,t) = \sum_n 2^{-n} d(z_n(s), z_n(t))$.

1. *Verify that* ρ *is indeed a metric.*
2. *Let* s_n *be a sequence in* S^∞, *that is,* s_n *is a sequence of sequences. Show that* $\rho(s_n, s) \to 0$ *iff for all* T, *there exists an* N *such that for all* $n \geq N$,

   $$(z_1(s_n), \ldots, z_T(s_n)) = (z_1(s), \ldots, z_T(s)).$$
3. *Let* s_n *be a sequence in* S^∞, *that is, a sequence of sequences. Show that* $\text{accum}(s_n)$ *is a non-empty subset of* S^∞ *so that* (S^∞, ρ) *is compact.*
4. *Show that every cylinder set is closed.* [*Since closed subsets of compact sets are compact, every cylinder set is in fact compact.*]

5. *Let μ be a finitely additive probability on $\mathcal{C}^\circ$ and let A_n be a sequence of cylinder sets with $A_n \downarrow \emptyset$. Using the finite intersection property, show that $\mu(A_n) \downarrow 0$, indeed, show the stronger result that there exists an N such that for all $n \geq N$, $\mu(A_n) = 0$.*

4. Probabilities on Fields and σ-Fields

We've already seen that $\{0,1\}^\infty$ is uncountable, it also looks a lot like the unit interval, $(0,1]$. For each $s \in \{0,1\}^\infty$, define $r_s = \sum_k z_k(s)/2^k \in [0,1]$. This maps $\{0,1\}^\infty$ onto $[0,1]$. For each $r \in (0,1]$, let s_r be the non-terminating binary expansion of r. This maps $(0,1]$ onto $\{0,1\}^\infty$.

This is meant to make it look reasonable to hope that we can simultaneously construct a model for drawing a point in the unit interval and drawing an infinite sequence of random variables. Discrete probabilities are just not enough to help us with the limit constructions we want, so we're going to develop a theory that allows us talk about probabilities on these uncountable spaces. We'll also see that finitely additive probabilities are also not enough and we'll develop countably additive probabilities.

4.1. Finitely additive probabilities on fields are a lot, but not quite enough.

This part is closely based on [3, Section 1, Ch. 1], which you should read. Let $\mathcal{B}^\circ$ be the empty set plus the collection of subsets of $(0,1]$ of the form $\cup_{k=1}^K (a_k, b_k]$ where each $(a_k, b_k] \subset (0,1]$.

Homework 4.1. *$\mathcal{B}^\circ$ is a field, and every non-empty $B \in \mathcal{B}^\circ$ can be expressed as a finite union of* disjoint *sets $(a_k, b_k]$.*

Define $\lambda((a,b]) = b - a$. We'll go crazy trying to keep enough brackets around, the "correct" way to write the last really is "$\lambda((a,b]) = b - a$," but we'll give ourselves permission to write "$\lambda(a,b] = b - a$," and we won't even be embarassed.

For every $B = \cup_{k=1}^K (a_k, b_k]$ with disjoint $(a_k, b_k]$, define $\lambda(B) = \sum_k \lambda(a_k, b_k]$.

Homework 4.2. *λ is a finitely additive probability on $\mathcal{B}^\circ$.*

This λ can give rise to all of the μ_θ on $\{0,1\}^\infty$ that we saw above.

Given a $0 < \theta < 1$, the θ-split of an interval $(a, b]$ is the partition of $(a, b]$,

$$\{I^\theta_{1,(a,b]} = (a, a + \theta(b-a)], I^\theta_{2,(a,b]} = (a + \theta(b-a), b]\}.$$

The idea is to inductively θ-split $(0, 1]$ into a sequence of finer and finer little disjoint subintervals.

Let $\mathcal{I}_1 = \{I^\theta_{1,1}, I^\theta_{2,1}\}$ be the θ-split of $(0, 1]$. Given $\mathcal{I}^\theta_n$ containing 2^n disjoint intervals, $I^\theta_{k,n}$, $1 \leq k \leq 2^n$, let $\mathcal{I}^\theta_{n+1} = \{I^\theta_{k,n+1} : 1 \leq k \leq 2^{n+1}\}$ be the collection of 2^{n+1} disjoint intervals, numbered from left to right, of θ-splits of the $I^\theta_{k,n}$.

Notation switch: Since we're starting to do probability theory here, we'll start referring to the probability space, here $(0, 1]$, as Ω, and to points in Ω as ω's.

Now, for each $n \in \mathbb{N}$, define the $\mathcal{B}^\circ$ measurable function

$$X^\theta_n(\omega) = \begin{cases} 1 & \text{if} \quad s \in I^\theta_{k,n}, \ k \text{ odd} \\ 0 & \text{if} \quad s \in I^\theta_{k,n}, \ k \text{ even} \end{cases}$$

Homework 4.3. *The $(X^\theta_n)_{n\in\mathbb{N}}$ are independent.*

Homework 4.4. *For each $\epsilon > 0$ and any $\theta \in (0, 1)$, $\lim_n p^\theta_n(\epsilon) = 0$ where*

$$p^\theta_n(\epsilon) = P\left\{\omega : \left|\frac{1}{n}\sum_{t=1}^{n} X^\theta_t(\omega) - \theta\right| \geq \epsilon\right\}.$$

The previous result says that if n is large, it is unlikely that the average of the X^θ_t's, $t \leq n$, is very far from θ. It is a version of the weak law of large numbers. The strong law of large numbers is the statement that, outside of a set of ω having probability 0, $\lim_n \frac{1}{n}\sum_{t=1}^n X^\theta_t(\omega) = \theta$. This is a very different kind of statement, it rules out every ω having some infinite sequence of times, $T_n(\omega)$ with $\left|\frac{1}{T_n(\omega)}\sum_{t=1}^{T_n(\omega)} X^\theta_t(\omega) - \theta\right| > \epsilon$. If the T_n were arranged to become sparser and sparser as n grows larger, this could still be consistent with the $\lim_n p^\theta_n(\epsilon) = 0$ condition just given.

Before going any further, let's look carefully at the set of ω we are talking about. For any ω and any T,

$$\lim_n \frac{1}{n}\sum_{t=1}^{n} X^\theta_t(\omega) = \lim_n \frac{1}{n}\sum_{t=T+1}^{n} X^\theta_t(\omega).$$

In $\{0,1\}^\infty$, this means that information about ω contained in any $\mathcal{C}_T^\circ$ is of no use in figuring out whether or not ω belongs to the set for which $\lim_n \frac{1}{n}\sum_{t=1}^n X_t^\theta(\omega) = \theta$. In Ω, this means that finite subdivisions of $(0,1]$ contained in $\mathcal{B}^\circ$ are insufficient to answer the kind of limit questions we'd like to answer.

What we need to do then, is to extend λ from $\mathcal{B}^\circ$ to a class of sets significantly larger than $\mathcal{B}^\circ$ that it contains the limit events we care about, and then, with that extension, still denoted by λ, show that

$$\lambda\left\{\omega : \lim_n \frac{1}{n}\sum_{t=1}^n X_t^\theta(\omega) = \theta\right\} = 1.$$

The class of sets "significantly larger" than $\mathcal{B}^\circ$ is called a σ-field. It is a field that has been closed, or completed, under countable limit operations. There is a useful intuitive analogy to the metric completion theorem, which adds new points for each of the non-convergent Cauchy sequences. The σ-field adds new sets for each of the non-convergent Cauchy sequences of sets.

Homework 4.5. *Do one of the following two:*

1. *Show that the complement of the set of* ω *such that* $\lim_n \frac{1}{n}\sum_{t=1}^n X_t^\theta(\omega) = \theta$ *is negligible.*
2. *Do any* 4 *problems from the end of* [3, Ch. 1, §1].

4.2. **The basics of σ-fields.** Recall that $\mathcal{F}^\circ$ is a field if

1. $S^\infty, \emptyset \in \mathcal{F}^\circ$,
2. if $A \in \mathcal{F}^\circ$, then $A^c \in \mathcal{F}^\circ$,
3. if $(A_m)_{m=1}^M \subset \mathcal{F}^\circ$, then $\cap_{m=1}^M A_m \in \mathcal{F}^\circ$.

Since $(\cap_m A_m)^c = \cup_m A_m^c$ (and you should check this) and fields contains the complements of all of their elements, we can replace "$\cap_{m=1}^M A_m \in \mathcal{F}^\circ$" by "$\cup_{m=1}^M A_m \in \mathcal{F}^\circ$" in the third line above.

Definition 4.1. *A class* $\mathcal{F}$ *of subsets of a set* Ω *is a* σ*-field if*

1. $S^\infty, \emptyset \in \mathcal{F}$,

2. *if* $A \in \mathcal{F}$*, then* $A^c = \in \mathcal{F}$*,*
3. *if* $(A_m)_{m\in\mathbb{N}} \subset \mathcal{F}$*, then* $\cap_{m\in\mathbb{N}} A_m \in \mathcal{F}$*.*

$(\cap_m A_m)^c = \cup_m A_m^c$ implies that we can replace "$\cap_{m\in\mathbb{N}} A_m \in \mathcal{F}$" by "$\cup_{m\in\mathbb{N}} A_m \in \mathcal{F}$" in the third line.

Verbally, a σ-field is a field that is closed under countable unions and intersections. If $A_n \subset A_{n+1}$ for all $n \in \mathbb{N}$, then we write $A_n \uparrow A$ where $A = \cup_n A_n$. In much the same way, if $A_n \supset A_{n+1}$ for all $n \in \mathbb{N}$, then we write $A_n \downarrow A$ where $A = \cap_n A_n$. If $\mathcal{F}$ is a field, then being closed under these two monotonic operations is the same as being a σ-field.

Lemma 4.2. *If $\mathcal{F}$ is a field, then $\mathcal{F}$ is a σ-field iff it is closed under monotone unions iff it is closed under monotone intersections.*

Proof: If $\mathcal{F}$ is a σ-field, then it is closed under all countable unions and intersections, whether or not they are monotonic. Suppose that $\mathcal{F}$ is a field that is closed under monotonic unions and let (A_n) be an arbitrary sequence of sets in $\mathcal{F}$, nested or not. Define $B_n = \cup_{m=1}^n A_m$. Since $\mathcal{F}$ is a field, each $B_n \in \mathcal{F}$, and $B_n \subset B_{n+1}$, so that $\cup_n B_n \in \mathcal{F}$. But $\cup_n B_n = \cup_n A_n$. The proof for intersections replaces each "$\cup$" by "$\cap$," and replaces "$B_n \subset B_{n+1}$" with "$B_n \supset B_{n+1}$." ∎

Factoids:

1. 2^Ω is a σ-field, it is the largest possible.
2. $\{\emptyset, \Omega\}$ is a σ-field, it is the smallest possible.
3. If each $\mathcal{F}_\alpha \subset 2^\Omega$ is a σ-field, then $\cap_\alpha \mathcal{F}_\alpha$ is a σ-field.
4. If $\mathcal{A} \subset 2^\Omega$, then $\sigma(\mathcal{A}) := \cap\{\mathcal{F} : \mathcal{F}$ is a σ-field, $\mathcal{A} \subset \mathcal{F}\}$ is the smallest σ-field containing $\mathcal{A}$. It is called the **σ-field generated by** $\mathcal{A}$. It is denoted $\sigma(\mathcal{A})$.

Of particular interest for us is the σ-field $\mathcal{B} := \sigma(\mathcal{B}^\circ)$. $\mathcal{B}$ is called the **Borel** σ-field in honor of Emile Borel who created a great deal of the mathematics we are studying.

There are two kinds of limit operations for sequences of sets that we will use fairly regularly. Let $(A_n)_{n\in\mathbb{N}} \subset \mathcal{F}$, $\mathcal{F}$ a σ-field. The set of points that are in all but at most finitely many of the A_n is called "$[A_n$ a.a.]," where "a.a." stands for "almost always." The set of points that are in infinitely many of the A_n is called "[A_n i.o.]," where "i.o." stands for "infinitely often."

There is a close connection to the ideas of $\liminf_n r_n$ and $\limsup_n r_n$ when r_n is a sequence in $\mathbb{R}$. We will use these notions often. We start with

Lemma 4.3. *If r_n is a bounded, monotonically increasing sequence (i.e. $r_n \leq r_{n+1}$), then $\lim_n r_n$ exists and is equal to $\sup\{r_n : n \in \mathbb{N}\}$.*

Proof: Since $\{r_n : n \in \mathbb{N}\}$ is a bounded set, it has a supremum, call it r. By the definition of the supremum, for all $\epsilon > 0$, there exists an $r_{N_\epsilon} \in \{r_n : n \in \mathbb{N}\}$ such that $r_{N_\epsilon} > r - \epsilon$. Since the sequence is monotonically increasing, for all $n \geq N_\epsilon$, $r_n > r - \epsilon$. Since all of the r_n are less than or equal to r (by the definition of a supremum), for all $n \geq N_\epsilon$, $|r_n - r| < \epsilon$, so that $r_n \to r$. ∎

If r_n is a bounded, monotonically **de**creasising sequence, we can replace "sup" by "inf" in Lemma 4.3. This leaves us ready for

Definition 4.4. *For a bounded sequence r_n,*

$$\limsup_n r_n := \lim_m \sup\{r_n : n \geq m\}, \quad \textit{and} \quad \liminf_n r_n := \lim_m \inf\{r_n : n \geq m\}.$$

Let $s_m = \sup\{r_n : n \geq m\}$, and note that s_m is monotonically decreasing. Therefore, by Lemma 4.3, it has a limit, specifically $\inf\{s_m : m \in \mathbb{N}\}$. Turning things around, let $t_m = \sup\{r_n : n \geq m\}$, and note that t_m is monotonically increasing. Therefore, by Lemma 4.3, it has a limit, specifically $\sup\{t_m : m \in \mathbb{N}\}$.

By the way, the last paragraph shows that we could just as well have used "$\inf\sup_n$" for "$\limsup_n$" and "$\sup\inf_n$" for and "$\liminf_n$."

To get to the connection to sets and the ideas of a.a. and i.o., let r_n be a bounded sequence, and define $A_n = (-\infty, r_n|$ (which means that I am identifying, i.e. declaring equivalent, the interval $(-\infty, r_n]$ and the interval $(-\infty, r_n)$, though this identification is just an emphemeral thing).

Homework 4.6. *Let r_n be a bounded sequence, and define $A_n = (-\infty, r_n| \subset \mathbb{R}$.*

1. $(-\infty, \liminf_n r_n| = \cup_m \cap_{n \geq m} A_n$*,*
2. $(-\infty, \limsup_n r_n| = \cap_m \cup_{n \geq m} A_n$*, and*

3. $(-\infty, \liminf_n r_n] \subset (-\infty, \limsup_n r_n]$.

More generally,

Homework 4.7. *For any sequence A_n of subsets of a non-empty Ω,*

1. $[A_n \text{ a.a.}] = \cup_m \cap_{n \geq m} A_n$.
2. $[A_n \text{ i.o.}] = \cap_m \cup_{n \geq m} A_n$.
3. $[A_n \text{ a.a.}] \subset [A_n \text{ i.o.}]$.

Sometimes $[A_n \text{ a.a.}]$ is called $\liminf_n A_n$ and $[A_n \text{ i.o.}]$ is called $\limsup_n A_n$. This should now make sense.

Homework 4.8. *Do problems* [3, Ch. 1, §2, 1, 4, 11]. *These are about the relation between the maxima and minima of indicator functions and unions and intersections, filtrations, and separable σ-fields respectively.*

Homework 4.9. *Do problems* [3, Ch. 1, §4, 1, 2, 5]. *These are about the relation between the* lim inf*'s and* lim sup*'s of sequences of indicator functions and* $[A_n \text{ a.a.}]$ *and* $[A_n \text{ i.o.}]$*, about properties of* $[A_n \text{ a.a.}]$ *and* $[A_n \text{ i.o.}]$*, and about convergence of sets in the sense that $P(A_n \Delta A) \to 0$ respectively.*

If the A_n belong to a σ-field $\mathcal{F}$, then each $B_m = \cap_{n \geq m} A_n \in \mathcal{F}$, implying that $\cup_m B_m \in \mathcal{F}$, so that $[A_n \text{ a.a.}] \in \mathcal{F}$.

Homework 4.10. *If $(A_n)_{n \in \mathbb{N}} \subset \mathcal{F}$ and $\mathcal{F}$ is a σ-field, then $[A_n \textit{ i.o.}] \in \mathcal{F}$.*

Of particular interest to us right now is the case where the $A_n \in \mathcal{B}^\circ$ and $\mathcal{B}$ is $\sigma(\mathcal{B}^\circ)$. The σ-field $\mathcal{B}$ is called the **Borel** σ-field in honor of Emile Borel. More factoids:

1. $A^\theta = \{\omega : \lim_n \frac{1}{n} \sum_{t=1}^n X_t^\theta(\omega) = \theta\} \in \mathcal{B}$.
2. $A^{\mathfrak{C}} = \{\omega : (\frac{1}{n} \sum_{t=1}^n X_t^\theta(\omega))_{n \in \mathbb{N}}$ is a Cauchy sequence $\} \in \mathcal{B}$.
3. $A^{\liminf} = \{\omega : \liminf_n (\frac{1}{n} \sum_{t=1}^n X_t^\theta(\omega))_{n \in \mathbb{N}}$ exists in $\mathbb{R}\} \in \mathcal{B}$.
4. $A^{\limsup} = \{\omega : \limsup_n (\frac{1}{n} \sum_{t=1}^n X_t^\theta(\omega))_{n \in \mathbb{N}}$ exists in $\mathbb{R}\} \in \mathcal{B}$.
5. $A^{\mathfrak{C}} \subset A^{\liminf} \cap A^{\limsup}$.

6. All of the above continue to be true if we replace $\frac{1}{n}\sum_{t=1}^{n} X_t^\theta(\omega)$ by an arbitrary sequence of functions $f_n(\omega)$ where $f_n(\omega)$ depends only on ω through the values of the first n X_t's. [This uses the special structure of $\{0,1\}^\infty$ a bit more than the previous factoids.]

So, we've got the probability λ defined on $\mathcal{B}^\circ$ and we would like to know $\lambda(E)$ for all of these $E \in \mathcal{B} = \sigma(\mathcal{B}^\circ)$. This requires extending λ from its domain, $\mathcal{B}^\circ$, to the larger domain $\mathcal{B}$. Ad astra.

4.3. **Extension of probabilities.** An essential result for limit theorems in probability theory is that every countably additive probability on a field $\mathcal{F}^\circ$ has a unique extension to $\mathcal{F} = \sigma(\mathcal{F}^\circ)$. Making this look reasonable by using the metric completion theorem is the aim of the present subsection.

Recall that a probability P on $\mathcal{F}^\circ$ is **countably additive** if $P(\Omega) = 1$, for any disjoint collection $\{A_1, \ldots, A_M\} \subset \mathcal{F}^\circ$, $P(\cup_m A_m) = \sum_m P(A_m)$, and if A_n is a sequence in $\mathcal{F}^\circ$ with $A_n \downarrow \emptyset$, then $\lim_n P(A_n) = 0$.

Definition 4.5. *A* **pseudo-metric** *on a set X is a function $d : X \times X \to \mathbb{R}_+$ such that*

1. *$d(x, y) = d(y, x)$, and*
2. *$d(x, y) + d(y, z) \geq d(x, z)$.*

Define $x \sim_d y$ if $d(x, y) = 0$. This is an equivalence relation, d defines a metric on the set of $\sim_d$ equivalence classes.

The function $d_P(A, B) = P(A\Delta B)$ is a pseudo-metric on $\mathcal{F}^\circ$. By way of a parallel, think of a utility function $u : \mathbb{R}_+^L \to \mathbb{R}$. We can define $d_u(x, y) = |u(x) - u(y)|$ to be the utility distance between x and y. The indifference surfaces are exactly the d_u-equivalence classes, and d_u measures the (utility) distance between indifference curves. Just as the consumer is indifferent between some very different points x and y, we will be indifferent between sets of points that differ only by a set of probability 0, even if that set having probability 0 is contains "many" points. To be quite explicit, we will not distinguish between sets A and B such that $d_P(A, B) = P(A \Delta B) = 0$.

The essential idea is to complete the pseudo-metric space $(\mathcal{F}^\circ, d_P)$, to discover that this is the pseudo-metric space $(\mathcal{F}, d_P)$, and to note that $P(E) = d_P(E, \emptyset)$ extends P from $\mathcal{F}^\circ$ to $\mathcal{F}$.

The proof of the following Theorem uses a number of principles and Lemmas that are important in their own right. Of these, the "good sets" principle and the first Borel-Cantelli Lemma will be seen most often in the future. Remember that Lemma 4.2 told us that fields that are also closed under monotonic unions or intersections are σ-fields.

Theorem 4.6. *If P is countably additive on $\mathcal{F}$, then the pseudo-metric space $(\mathcal{F}, d_P)$ is complete. If $\mathcal{F}^\circ$ is a field generating $\mathcal{F}$, then $\mathcal{F}^\circ$ is d_P-dense in $\mathcal{F}$.*

Proof: There are two parts to the proof, denseness and completeness.
Denseness: Let us first show that $\mathcal{F}^\circ$ is d_P-dense in $\mathcal{F}$. This part of the proof uses the "good sets" principle. One names a class of sets having the property you want, show that it contains a generating class of sets, and that it's a field closed under monotonic unions or intersections. This means that the class of good sets is a σ-field containing a generating class, i.e. it's contains the σ-field we're interested in.

Let $\mathcal{G}$ denote the class of "good sets" for this proof, that is,

$$\mathcal{G} = \{E \in \mathcal{F} : (\forall \epsilon > 0)(\exists E^\epsilon \in \mathcal{F}^\circ)[d(E, E^\epsilon) < \epsilon]\}.$$

The three steps are to show that $\mathcal{G}$ contains a generating class, is a field, and is closed under monotonic unions.

$\mathcal{G}$ contains $\mathcal{F}^\circ$:

There's not much to prove here, if $E \in \mathcal{F}^\circ$, take $E^\epsilon = E$.

$\mathcal{G}$ is a field:

1. $\emptyset, \Omega \in \mathcal{G}$ because $\emptyset, \Omega \in \mathcal{F}^\circ$.
2. Suppose that $E \in \mathcal{G}$. Pick $\epsilon > 0$ and E^ϵ such that $d_P(E, E^\epsilon) < \epsilon$. For any $A, B \in \mathcal{F}$, $d_P(A, B) = d_P(A^c, B^c)$ so that the complement of E^ϵ ϵ-approximates E^c, so that $E^c \in \mathcal{G}$.
3. Suppose $(A_m)_{m=1}^M \subset \mathcal{G}$. Pick arbitrary $\epsilon > 0$ and $E_m \in \mathcal{F}^\circ$ such that $d_P(A_m, E_m) < \epsilon/M$. Because $(\cup_{m=1}^M A_m)\Delta(\cup_{m=1}^M E_m) \subset \cup_{m=1}^M (A_m \Delta E_m)$ and $P(\cup_{m=1}^M (A_m \Delta E_m)) \leq \sum_{m=1}^M \epsilon/M = \epsilon$, $d_P(\cup_{m=1}^M A_m, \cup_{m=1}^M E_m) < \epsilon$.

$\mathcal{G}$ **is closed under monotonic unions**:

Let $A_n \uparrow A$, $A_n \in \mathcal{G}$, we need to show that $A \in \mathcal{G}$. For this purpose, pick an arbitrary $\epsilon > 0$. We must show that there exists a set in $\mathcal{G}$ at d_P-distance less than ϵ from A. The sequence $A \setminus A_n \downarrow 0$ so that $P(A_n) \uparrow P(A)$. Therefore we can pick N such that for all $n \geq N$, $|P(A) - P(A_n)| < \epsilon/2$. Because $A_n \subset A$, $d_P(A, A_n) = |P(A) - P(A_n)| < \epsilon/2$. Since $A_N \in \mathcal{G}$, we can pick an $A_N^{\epsilon/2} \in \mathcal{F}^\circ$ such that $d_P(A_N^{\epsilon/2}, A_N) < \epsilon/2$. By the triangle inequality, $d_P(A, A_N^{\epsilon/2}) \leq d_P(A, A_N) + d_P(A_N, A_N^{\epsilon/2}) < \epsilon/2 + \epsilon/2 = \epsilon$. Since $A_N^{\epsilon/2} \in \mathcal{F}^\circ$, $A \in \mathcal{G}$.

That completes the denseness part of the proof.

Completeness: Let A_n be a Cauchy sequence in $\mathcal{F}$. We will take a subsequence A_{n_k} such that $\lim_k d_p(A, A_{n_k}) = 0$ for some $A \in \mathcal{F}$. By the triangle inequality, $d_P(A_n, A) \to 0$ because A_n is Cauchy.

First, the inductive construction of the subsequence: Pick n_1 such that for all $n, m \geq n_1$, $d_P(A_n, A_m) < 2^{-1}$. Given that n_{k-1} has been picked, pick $n_k > n_{k-1}$ such that for all $n, m \geq n_k$, $d_P(A_n, A_m) < 2^{-k}$. Note that $\sum_k d_P(A_{n_k}, A_{n_{k+1}}) = \sum_k P(A_{n_k} \Delta A_{n_{k+1}}) < \sum_k 2^{-k} < \infty$. We will use the following result, which is quite important in its own right (despite the fact that it is so easy to prove).

Lemma 4.7 (Borel-Cantelli)**.** *If P is countably additive and A_n is a sequence in $\mathcal{F}$ such that $\sum_n P(A_n) < \infty$, then $P([A_n \text{ i.o.}]) = 0$.*

Proof: For every m, $[A_n \text{ i.o.}] \subset \cup_{n \geq m} A_n$ so that $P([A_n \text{ i.o.}]) \leq P(\cup_{n \geq m} A_n) \leq \sum_{n \geq m} P(A_n)$. Since $\sum_n P(A_n) < \infty$, $\sum_{n \geq m} P(A_n) \downarrow 0$ as $m \uparrow \infty$. ∎

Let us relabel each A_{n_k} as A_k so that we don't have to keep track of two levels of subscripts. From the Borel-Cantelli Lemma and the construction, we know that $P[A_k \Delta A_{k+1} \text{ i.o.}] = 0$.

Second, we are going to show that $P([A_k \text{ i.o.}] \setminus [A_k \text{ a.a.}]) = 0$. Since $[A_k \text{ a.a.}] \subset [A_k \text{ i.o.}]$, this means that $d_P([A_k \text{ a.a.}], [A_k \text{ i.o.}]) = 0$, i.e. that the two sets are in the same d_P-equivalence class. The proof that $P([A_k \text{ i.o.}] \setminus [A_k \text{ a.a.}]) = 0$ consists of showing that

$$([A_k \text{ i.o.}] \setminus [A_k \text{ a.a.}]) \subset [A_k \Delta A_{k+1} \text{ i.o.}].$$

Pick an arbitrary $\omega \in ([A_k \text{ i.o.}] \setminus [A_k \text{ a.a.}])$. Since $\omega \notin [A_k \text{ a.a.}]$, we know that $\omega \in [A_k^c \text{ i.o.}]$. Therefore, $\omega \in [A_k \text{ i.o.}]$ and $\omega \in [A_k^c \text{ i.o.}]$. This means that for infinitely many k, either $\omega \in A_k \setminus A_{k+1}$ or $\omega \in A_{k+1} \setminus A_k$. This is exactly the same as saying that $\omega \in [A_k \Delta A_{k+1} \text{ i.o.}]$.

Finally, we need to show that $d_P(A_K, A) \to 0$. (By the way, in doing this, we'll be doing most of the homework problem [3, Ch. 1, §4, 5].) For each K, let

$$B_K = \cap_{k \geq K} A_k, \text{ and } C_K = \cup_{k \geq K} A_k.$$

By the definitions of $[A_k \text{ a.a.}]$ and $[A_k \text{ i.o.}]$, we have, for all K,

$$B_K \subset [A_k \text{ a.a.}] \subset [A_k \text{ i.o.}] \subset C_K,$$

and

$$B_K \uparrow [A_k \text{ a.a.}] \text{ while } C_K \downarrow [A_k \text{ i.o.}].$$

By countable addtivity, this means that

$$P(B_k) \uparrow P[A_k \text{ a.a.}] \text{ and } P(C_K) \downarrow P[A_k \text{ i.o.}].$$

Since we have established that $P[A_k \text{ a.a.}] = P[A_k \text{ i.o.}]$, this means that $|P(C_K) - P(B_K)| \downarrow 0$. Now, $d_P(A_K, A) = P(A_K \setminus A) + P(A \setminus A_K)$. The proof is complete once we notice that

$$(A_K \setminus A) \cup (A \setminus A_K) \subset (C_K \setminus B_K).$$

To be completely explicit, we therefore have $d_P(A_K, A) \leq P(C_K \setminus B_K) \downarrow 0$. ∎

This Theorem means that, if we have already extended P to $\mathcal{F}$, then any field $\mathcal{F}^\circ$ with $\mathcal{F} = \sigma(\mathcal{F}^\circ)$ is d_P-dense in $\mathcal{F}$, and $\mathcal{F}$ is the metric completion of $\mathcal{F}^\circ$. Ideally, the next set of arguments start with the metric space $(\mathcal{F}^\circ, d_P)$, P countably additive, sets $(\widehat{\mathcal{F}^\circ}, \widehat{d_P})$ as its metric completion, and then identifies the "points" in $\widehat{\mathcal{F}^\circ} \setminus \mathcal{F}^\circ$ as d_P-equivalence classes of elements of $\mathcal{F} = \sigma(\mathcal{F}^\circ)$ that are not already contained in $\mathcal{F}^\circ$. It certainly seems plausible that this is doable, and it is. Unfortunately, the only way that I have found to do it is tricky beyond its worth.[4] So, I will (a bit shamefacedly) simply state the Theorem, a very good proof is in [3, Ch. 1, §3].

Theorem 4.8. *Every countably additive P on a field $\mathcal{F}^\circ$ has a unique, countably additive extension to $\mathcal{F} = \sigma(\mathcal{F}^\circ)$.*

4.4. **The Tail σ-field and Kolmogorov's 0-1 Law.** Fix a probability space $(\Omega, \mathcal{F}, P)$, $\mathcal{F}$ a σ-field and P a countably additive probability on $\mathcal{F}$. If $\mathcal{F}_\alpha$ and $\mathcal{F}$ are σ-fields and $\mathcal{F}_\alpha \subset \mathcal{F}$, then we say that $\mathcal{F}_\alpha$ is a **sub-σ-field of $\mathcal{F}$**.

Definition 4.9. *A collection $\{\mathcal{C}_\alpha : \alpha \in A\}$ of subsets of $\mathcal{F}$ is* **independent** *if for any finite $A' \subset A$ and any choices $E_\alpha \in \mathcal{C}_\alpha$, $P(\cap_{\alpha \in A'} E_\alpha) = \Pi_{\alpha \in A'} P(E_\alpha)$.*

[4]If I ever knew an easy version of the argument, I have forgotten it. The only one I can presently find passes through a transfinite induction argument. The $\pi - \lambda$ Theorem and the Monotone Class Theorem used in most proofs are clever ways to avoid doing transfinite induction.

You should learn (or have learned) examples showing that pairwise independence is weaker than independence.

Theorem 4.10. *If the collection $\{\mathcal{C}_\alpha : \alpha \in A\}$ of subsets of $\mathcal{F}$ is independent and each $\mathcal{C}_\alpha$ is closed under finite intersection, then the collection $\{\sigma(\mathcal{C}_\alpha) : \alpha \in A\}$ is independent.*

Before proving Theorem 4.10, we'll prove the (very useful) $\pi - \lambda$ theorem.

Definition 4.11. *A class $\mathcal{L}$ of subsets of Ω is called a λ system (or "une classe σ-additive d'ensembles" if you follow the French tradition) if*

1. $\Omega \in \mathcal{L}$,
2. *$\mathcal{L}$ is closed under disjoint unions,*
3. *$\mathcal{L}$ is closed under proper differences, i.e. if $E_1, E_2 \in \mathcal{L}$ and $E_1 \subset E_2$, then $E_2 \setminus E_1 \in \mathcal{L}$, and*
4. *if E_n is a sequence in $\mathcal{L}$ and $E_n \uparrow E$, then $E \in \mathcal{L}$.*

Notice that any σ-field is a λ system. There is another parallel: The intersection of an arbitrary collection of λ systems is again a λ system, and 2^Ω is a λ system. This shows that any class $\mathcal{C}$ of subsets of Ω is contained in a smallest λ system, called the λ system generated by $\mathcal{C}$ and written $\mathcal{L}(\mathcal{C})$.

A class $\mathcal{P}$ of subsets of Ω is called a π system if it is closed under finite intersection.

Theorem 4.12 (π-λ)**.** *If $\mathcal{P}$ is a π system, then $\mathcal{L} = \mathcal{L}(\mathcal{P}) = \sigma(\mathcal{P})$.*

Proof: Since $\mathcal{L} \subset \sigma(\mathcal{P})$, it is enough to show that $\mathcal{L}$ is a σ-field. We know that $\Omega \in \mathcal{L}$. Since $\Omega \in \mathcal{L}$ and $\mathcal{L}$ is closed under proper differences, $E \in \mathcal{L}$ implies $(\Omega \setminus E) = E^c \in \mathcal{L}$. Since $\mathcal{L}$ is a monotone class, all that is left is to show that $\mathcal{L}$ is closed under intersection. This involves a clever bit of dodging around.

Let

$$\mathcal{G}_1 = \{E \in \mathcal{L} : E \cap F \in \mathcal{L} \text{ for all } F \in \mathcal{P}\},$$

and let

$$\mathcal{G}_2 = \{E \in \mathcal{L} : E \cap F \in \mathcal{L} \text{ for all } F \in \mathcal{L}\}.$$

Note that if $\mathcal{G}_2 = \mathcal{L}$, then $\mathcal{L}$ is closed under finite intersection.

First we will verify that $\mathcal{G}_1$ is a λ system containing $\mathcal{P}$, which tells us that $\mathcal{G}_1 = \mathcal{L}$. Then, we note that $\mathcal{G}_1 = \mathcal{L}$ implies that $\mathcal{P} \subset \mathcal{G}_2$. Finally, we verify that $\mathcal{G}_2$ is also a λ system, so that $\mathcal{G}_2 = \mathcal{L}$.

$\mathcal{G}_1$ is a λ system containing $\mathcal{P}$: $\mathcal{G}_1$ contains $\mathcal{P}$ because $\mathcal{P}$ is closed under finite intersection. It contains Ω by inspection. If E_1 and E_2 are disjoint elements of $\mathcal{G}_1$, then $E_1 \cap F \in \mathcal{L}$ and $E_2 \cap F \in \mathcal{L}$ for all $F \in \mathcal{P}$. Since $\mathcal{L}$ is closed under disjoint unions and $E_1 \cap F$ and $E_2 \cap F$ are disjoint and belong to $\mathcal{L}$, $(E_1 \cap F) \cup (E_2 \cap F) = (E_1 \cup E_2) \cap F \in \mathcal{L}$ for all $F \in \mathcal{P}$. Proper differences and monotonic increasing sequences are checked by the same logic.

$\mathcal{G}_2$ is a λ system containing $\mathcal{P}$: From the previous step, $\mathcal{P} \subset G_2$. Verifying that $\mathcal{G}_2$ is a λ system is direct. ∎

If the collection $\{\mathcal{C}_\alpha : \alpha \in A\}$ of subsets of $\mathcal{F}$ is independent and each $\mathcal{C}_\alpha$ is closed under finite intersection, then the collection $\{\sigma(\mathcal{C}_\alpha) : \alpha \in A\}$ is independent.

Proof of Theorem 4.10: For any α, let $\mathcal{D}_\alpha$ be the set of $E \in \sigma(\mathcal{C}_\alpha)$ with the property that for any finite collection $E_{\alpha'} \in C_{\alpha'}$ indexed by distinct α',

$$P(E \cap \bigcap_{\alpha'} E_{\alpha'}) = P(E) \times \Pi_{\alpha'} P(E_{\alpha'}).$$

Each $C_{\alpha'}$ is a π system, and it is pretty easy to show that $\mathcal{D}_\alpha$ is a λ system because the $\mathcal{C}_{\alpha'}$ are closed under finite intersection. From this (and the π-λ Theorem) we conclude that $\mathcal{D}_\alpha = \sigma(\mathcal{C}_\alpha)$ for any α. This means that the collection $\{\sigma(\mathcal{C}_\alpha), \{C_{\alpha'} : \alpha' \neq \alpha\}\}$ is independent. Reapplying this theorem as often as needed (remembering that each $\sigma(\mathcal{C}_\alpha)$ is closed under finite intersection), for any finite $B \subset A$, the collection $\{\{\sigma(\mathcal{C}_\alpha) : \alpha \in B\}, \{C_{\alpha'} : \alpha' \notin B\}\}$ is independent. Going back to look at the definition of independence, we see that we're done. ∎

Definition 4.13. *Let $\{\mathcal{B}_n : n \in \mathbb{N}\}$ be a collection of sub-σ-fields of $\mathcal{F} := \sigma(\mathcal{B}_n : n \in \mathbb{N})$, let $\mathcal{F}_n = \sigma\{\mathcal{B}_m : n \leq m\}$, and let $\mathcal{F}_{n+} = \sigma\{\mathcal{B}_m : n \geq m\}$. The σ-field $\mathcal{F}_\tau := \cap_n \mathcal{F}_{n+}$ is called the* **tail σ-field** *or the* **tail σ-field generated by** *$\{\mathcal{B}_n : n \in \mathbb{N}\}$.*

Theorem 4.14 (Kolmogorov's 0-1 Law)**.** *If the $\mathcal{B}_n$ are independent and $A \in \mathcal{F}_\tau$, then $P(A) = 0$ or $P(A) = 1$.*

Proof: Applying Theorem 4.10, for each $n \in \mathbb{N}$, $\mathcal{F}_n$ is independent of $\mathcal{F}_{n+}$. Since $\mathcal{F}_\tau \subset \mathcal{F}_{n+}$, for each $n \in \mathbb{N}$, $\mathcal{F}_n$ is independent of $\mathcal{F}_\tau$. Applying Theorem 4.10 again, $\mathcal{F} = \sigma(\mathcal{F}_n : n \in \mathbb{N})$ is independent of $\mathcal{F}_\tau$. Now, pick an arbitrary $A \in \mathcal{F}_\tau$. Since

$\mathcal{F}_\tau \subset \mathcal{F}$, we know that A is independent of itself so that $P(A) \cdot P(A) = P(A \cap A) = P(A)$. The only numbers satisfying $a^2 = a$ are 0 and 1. ∎

4.5. **Measurability and the importance of the tail σ-field.** Fix a probability space $(\Omega, \mathcal{F}, P)$ and a complete separable metric space (csm) (M, d). Let $\mathcal{M}$ denote the Borel σ-field on M, that is, the σ-field generated by the open balls $B(x, \epsilon)$. [Warning: if you ever end up interested in a non-separable metric space, this is not the definition of the Borel σ-field, [21] shows that the distinction between this definition and the other one is useful for stochastic process theory.] The following is important WAY beyond what you might guess from the simplicity of the definition.

Definition 4.15. *A function $X : \Omega \to M$ is* **simple** *if X takes on only finitely many values. A simple function X is* **measurable** *if, for each point $x \in M$, $X^{-1}(m) \in \mathcal{F}$. More generally, a function X is* **measurable** *if there exists a sequence X_n of simple measurable functions such that $P\{\omega : X_n(\omega) \to X\} = 1$. A measurable function is also called a* **random variable**.

So, a measurable function is almost a simple measurable function.

If X_n is any sequence of simple measurable functions, then

$$C = \{\omega : X_n(\omega) \text{ converges }\} \in \mathcal{F}$$

by arguments we gave above (remember, (M, d) is complete so that convergent sequences are Cauchy sequences $\ldots$). Therefore, asking that $P\{\omega : X_n(\omega) \to X\} = 1$ is asking that $P(C) = 1$ and naming the function $X(\omega)$ as the limit of the $X_n(\omega)$ for each $\omega \in C$.

Definition 4.16. *For any sequence of measurable functions X, X_n, we say that X_n* **converges to X P-almost everywhere (a.e.)** *if $P\{X_n \to X\} = 1$.*

Homework 4.11. *If X, X_n is any sequence of random variables, then $\{X_n \to X\} \in \mathcal{F}$.*

Homework 4.12. *If X_n converges to X a.e., then for all $\epsilon > 0$,*

$$P\{\omega : d(X_n(\omega), X(\omega)) > \epsilon\} \to 0.$$

One reason that this definition is important is that a measurable X gives rise to a countably additive probability on $(M, \mathcal{M})$.

Lemma 4.17. *X is measurable if and only if $X^{-1}(A) \in \mathcal{F}$ for each $A \in \mathcal{M}$.*

If $X^{-1}(A) \in \mathcal{F}$ for each $A \in \mathcal{M}$, then we can define the $\mu = X(P)$ by $\mu_X(A) = P(X^{-1}(A))$. The measurable functions are exactly the functions that give rise to countably additive probabilities on their csm range spaces, exactly the ones for which we can assign a probability to the event that $X \in A$.

Homework 4.13. *Check that μ_X is countably additive.*

Proof of Lemma 4.17: Suppose that X is measurable and simple. Then it is easy. Now suppose that X is not simple. Let $\mathcal{G}$ be the class of sets $A \in \mathcal{M}$ such that $X^{-1}(A) \in \mathcal{F}$. Then show that $\mathcal{G}$ is a σ-field. Finally, show that $X^{-1}(B(x, \epsilon)) \in \mathcal{F}$ by showing that for all $\omega \in C$, $X^{-1}(B(x, \epsilon)) = [X_n^{-1}(B(x, \epsilon))$ a.a.].

Suppose that $X^{-1}(A) \in \mathcal{F}$ for each $A \in \mathcal{M}$. Follow your nose. ∎

This result motivates the general definition (useful for contexts when we don't have a complete separable metric space structure around).

Definition 4.18. *A function f from a measure space $(X, \mathcal{X})$ to another measure space $(Y, \mathcal{Y})$ is* **measurable** *if $f^{-1}(\mathcal{Y}) \subset \mathcal{X}$.*

Measurable functions of measurable functions are measurable.

Lemma 4.19. *If f is a measurable function from $(X, \mathcal{X})$ to space $(Y, \mathcal{Y})$ and g is a measurable function from $(Y, \mathcal{Y})$ to space $(Z, \mathcal{Z})$, then $f(g(x))$ is a measurable function from $(X, \mathcal{X})$ to $(Z, \mathcal{Z})$.*

We started with a σ-field and defined the set of measurable functions with respect to that σ-field. We can start with a measurable function, X, and define $\sigma(X) \subset \mathcal{F}$ to be $X^{-1}(\mathcal{M})$.

Definition 4.20. *If $\mathcal{G}$ is a sub-σ-field of $\mathcal{F}$, then X is $\mathcal{G}$-measurable if $\sigma(X) \subset \mathcal{G}$.*

We may later need one of the many results due to Doob: Y is $\sigma(X)$-measurable iff $Y = f(X)$ for some measurable f. If we need it, we'll prove it. Meanwhile,

Definition 4.21. *A collection of random variables* $(X_\alpha)_{\alpha\in A}$ *is* **independent** *if the collection of σ-fields* $(\sigma(X_\alpha))_{\alpha\in A}$ *is independent.*

Homework 4.14. *If* X_n *is a sequence of independent* $\mathbb{R}$*-valued random variables and* c_n *is a sequence of constants, then the following sets have probability either* 0 *or* 1 *:*

1. $\{c_n X_n \text{ is convergent}\}$,
2. $\{\sum_n |c_n X_n| < \infty\}$,
3. $\{\limsup_N \sum_{n=1}^N c_n X_n = \infty\}$, *and*
4. $\{\limsup_N c_N \cdot (\sum_{n=1}^N X_n) = 1\}$.

Homework 4.15. *Show that* $\sum_n \frac{1}{n} = \infty$. *If* X_n *is a sequence of independent random variables with* $P(X_n = 1) = P(X_n = -1) = \frac{1}{2}$, *find in* [3] *the result that* $P(R_n \text{ converges}) = 1$ *where the sequence of random variables* $R_N := \sum_{n=1}^N \frac{1}{n} X_n$.

The sequence $Y_n = \frac{1}{n} X_n$ is an example of what is called a martingale. We'll have occasion to talk about martingales later.

4.6. Detour #3: Failures of Countable Additivity and the Theory of Choice Under Uncertainty.

4.6.1. *Background.* Here is a sketch of a canonical probability on the integers that fails countable additivity, it is the "uniform" distribution. Any finitely additive probability on $\mathbb{N}$ is a function $P : 2^{\mathbb{N}} \to [0, 1]$. As such it can be represented as an infinitely long vector $(P(E))_{E\in 2^{\mathbb{N}}}$, this is a point in the infinite product space $\times_{E\in 2^{\mathbb{N}}}[0, 1]$. This is a really long vector.

Homework 4.16. $2^{\mathbb{N}}$ *is uncountable.*

Let P^n be a sequence of finitely additive probabilities. There is a very deep mathematical result (Alaoglu's Theorem) that says that any infinite set in $\times_{E\in 2^{\mathbb{N}}}[0, 1]$ has an accumulation point, P. Further, it says that

1. if $P_n(E)$ is convergent for some $E \in 2^{\mathbb{N}}$, then at any accumulation point, $P(E) = \lim_n P_n(E)$, and more generally,
2. if $f : [0, 1]^M \to \mathbb{R}$ is continuous, and $f(P_n(E_1), P_n(E_2), \ldots , P_n(E_M))$ is convergent, then at any accumulation point P,
$$f(P(E_1), \ldots , P(E_M)) = \lim_n f(P_n(E_1), P_n(E_2), \ldots , P_n(E_M)).$$

Homework 4.17. *Any accumulation point of a sequence of finitely additive probabilities must be finitely additive. [Hint: pick the right f above.]*

Let Λ be an accumulation point of the sequence Λ_n where Λ_n is the uniform distribution on $\{1, 2, \dots, n\}$.

Homework 4.18. *Show that*

1. *If E is finite, then $\Lambda(E) = 0$.*
2. $\Lambda(\{evens\}) = \frac{1}{2}$.
3. *Λ fails to be countably additive.*
4. *Λ is non-atomic – for any $\epsilon > 0$, it is possible to partition $\mathbb{N}$ into finitely many sets E_i with $\Lambda(E_i) < \epsilon$.*

For any bounded $\mathbb{R}$-valued function g on $\mathbb{N}$ is Λ-integrable, and the integral can be defined by

$$\int_{\mathbb{N}} g(n)\, d\Lambda(n) = \lim_{m\uparrow\infty} \sum_{i=-m2^m}^{+m2^m} \frac{i}{2^m} \Lambda\{g \in \left[\frac{i}{2^m}, \frac{i+1}{2^m}\right)\}. \tag{2}$$

Homework 4.19. *Suppose that two $\mathbb{R}$-valued functions f and g on $\mathbb{N}$ satisfy satisfy $f(m) > g(m) \geq 0$ for all $m \in \mathbb{N}$ and $\lim_{m\to\infty} f(m) = 0$. Then f and g are bounded and*

$$\int_{\mathbb{N}} f(m)\, d\Lambda(m) = \int_{\mathbb{N}} g(m)\, d\Lambda(m) = 0 \tag{3}$$

One generally avoids defining conditions by their failure, but ...

Definition 4.22. *A probability P* **fails conglomerability** *if there exists a countable partition $\pi = \{E_1, E_2, \dots\}$ of $\mathbb{N}$ some event $E \in 2^{\mathbb{N}}$, and constants $k_1 \leq k_2$ such that $k_1 \leq P(E|E_i) \leq k_2$ for each $E_i \in \pi$, yet $P(E) < k_1$ or $P(E) > k_2$.*

Failing conglomerability means that there is an event E, and a partition π with the property that, conditional on each and every event in π, the posterior probability of E is above (or below) the prior probability of E.

Theorem 4.23. *P is countably additive iff it is conglomerable.*

Homework 4.20. *Prove at least one direction of this Theorem.*

A simple version of Lebesgue's Dominated Convergence Theorem will be useful:

Homework 4.21. *Suppose that X_n is a sequence of random variables on a probability space $(\Omega, \mathcal{F}, P)$ with countably additive P and that the X_n are dominated in absolute value a.e., i.e. there exists some $M > 0$ such that for all n, $P\{|X_n| \leq M\} = 1$.*

1. *If $X_n \to X$ a.e., then $\int X_n\, dP \to \int X\, dP$.*
   This can also be written as

$$\lim_n \int X_n\, dP = \int \lim_n X_n\, dP,$$

that is, limit signs and integral signs can be interchanged when P is countably additive and the X_n are uniformly bounded. [The uniform boundedness condition can be relaxed in important ways.] The countable additivity cannot be relaxed at all.

2. *If P fails to be countably additive, then there exists a sequence of uniformly bounded random variables converging a.e. to some X with $\int X_n\, dP \not\to \int X\, dP$.*

To summarize, Dominated Convergence is equivalent to countable additivity.

4.6.2. *Savage preferences over acts and gambles.* For our present purposes, **acts** are functions from the measure space $(\mathbb{N}, 2^{\mathbb{N}})$ to a set of consequences C, always taken to be a csm, most often C taken to be a bounded interval in $\mathbb{R}$. The subjective probability P on $2^{\mathbb{N}}$ may vary, but will often be Λ.

Homework 4.22. *All acts are measurable.*

Under study are preferences (complete, transitive orderings) on the set of acts. Savage preferences, $\succeq$, over acts can be represented by a bounded utility function $u : C \to \mathbb{R}$ such that

$$[a_1 \succeq a_2] \Leftrightarrow \left[\int u(a_1(n))\, d\Lambda(n) \geq \int u(a_2(n))\, d\Lambda(n)\right].$$

The function u is call the **expected utility function**. Preferences over constant acts are particularly simple, if $a_1(n) \equiv c_1$ and $a_2(n) \equiv c_2$, then $a_1 \succeq a_2$ iff $u(c_1) \geq u(c_2)$.

We are going to assume, unless explicitly noted, that the preferences are non-trivial, i.e. there exists c_1 and c_2 such that $c_1 \succ c_2$, and that any Savage preferences are continuous, that is, u is a continuous function.

Definition 4.24. *Preferences over acts* **respect strict dominance** *if*

$$[(\forall n \in \mathbb{N})[a_1(n) \succ a_2(n)]] \Rightarrow [a_1 \succ a_2].$$

Savage preferences with finitely additive probabilities do <u>not</u> generally respect strict dominance.

Homework 4.23. *Let $\succeq_\Lambda$ be the Savage preferences over acts into the space of consequences $[-1, +1]$ be given by the subjective probability Λ and a continuous, strictly increasing utility function $u\colon [-1, +1] \to \mathbb{R}$. Let $\succeq_P$ be the Savage preferences with the same u and a countably additive subjective probability P. Suppose that $a_1(n) \downarrow 0$ and $a_1(n) > a_2(n) \geq 0$ so that a_1 strictly dominates a_2.*

1. $a_1 \sim_\Lambda a_2$.
2. $a_1 \succ_P a_2$.

A **money pump** is a sequence of acts that an agent would pay you to acquire with the unfortunate property that at the end of the process of taking them all, the agent would pay you to take them back. You get them coming and going, pumping money out of them. Money pumps exist when the subjective probabilities are not countably additive.

Some more terminology: Gambles are simple acts, that is, acts that take on only finitely many values, usually 2. Recall that for $A \in 2^{\mathbb{N}}$, $1_A(m)$, the indicator function of the set A, is the function taking on the value 1 if $m \in A$ and 0 if $m \notin A$.

Homework 4.24 (Adams)**.** *With the state space $\mathbb{N}$, let Q be the countably additive probability satisfying $Q\{n\} = 2^{-n}$. The subjective probability is $P = (Q + \Lambda)/2$ so that $P\{n\} = 2^{-(n+1)}$ and $\sum_{n=1}^{\infty} P\{n\} = \frac{1}{2} < P(\mathbb{N}) = 1$. The set of consequences is $[-1, +1]$, and the expected utility function is $U(x) = x$ so the agent is risk neutral. Fix some $r \in (\frac{1}{2}, 1)$. For each $n \in \mathbb{N}$, consider the gamble g_n that loses r if $B_n = \{n\}$ occurs, and that pays $2^{-(n+1)}$ no matter what occurs, that is,*

$$g_n(m) = 2^{-(n+1)} - r \cdot 1_{B_n}(m).$$

1. *Each g_n has a strictly positive expected value.*
2. *For all N, $\sum_{n=1}^{N+1} g_n \succ_P \sum_{n=1}^{N} g_n \succ_P 0$.*
3. $0 \succ_P \sum_{n=1}^{\infty} g_n$.
4. *For each N and m in the state space, let $X_N(m) = u(\sum_{n=1}^{N} g_n(m))$ and let $X(m) = u(\sum_{n=1}^{\infty} g_n(m))$. The sequence X, X_N of random variables is uniformly bounded. Show that for all m, $X_N(m) \to X(m)$, but that $\lim_N \int X_n \, dP \neq \int X \, dP$.*

The following money pump involves a countably infinite construction, but doesn't require countably many separate decisions. Part of the following problem involves figuring out what it means to prefer one act over another conditional on some event. It should be obvious to you if you think about the Bridge-Crossing Lemma.

Homework 4.25 (Dubins, then Seidenfeld and Schervish)**.** *Let $S = \cup\{(i,j)\colon\ i \in \mathbb{N},\ j = 0, 1\}$, so that S is the union of two copies of the integers, indexed by $j = 0$ or $j = 1$. The σ-field is 2^S. Let $E = \cup_i\{(i,1)\}$ be the event that $j = 1$, and for $i \in \mathbb{N}$. Let $E_i = \{(i,0),(i,1)\}$ so that $\pi = \{E_1, E_2, \dots\}$ is a partition of S. Conditional on E, suppose that $P(i,1) = 1/2(Q + \Lambda)(i)$ where Q and Λ are as in the previous problem. Conditional on E^c, suppose that $P = Q$.*

1. *For any $i \in \mathbb{N}$, $P(\{(i,0)\}) = \frac{1}{2} \cdot 2^{-i}$ and $P(\{(i,1)\}) = \frac{1}{4} \cdot 2^{-i}$.*
2. *For each E_i, $P(E|E_i) = \frac{1}{3}$ even though $P(E) = \frac{1}{2}$, so P is not conglomerable in π.*
3. $\sum_{E_i \in \pi} P(E_i) = \frac{3}{4} < 1$ *even though π is a partition.*
4. *Suppose that a_1 deliver a consequence worth 35 utils in all states while a_2 delivers a consequence worth 0 utils if E occurs and 60 utils if E does not occur. $a_2 \prec a_1$, but $a_1 \prec a_2$ given any E_i.*
5. *Let D_n be the complement of $\cup_{i=1}^{N} E_i$, and let $D = \cap_n D_n$. If P were countably additive, then $\lim_n \int 1_{D_n}(m)\, dP(m) = 1/4 > 0$ would imply that $P(D) = 1/4$ (this of Lebesgue's Dominated Convergence Theorem). However, the event D is the empty set, giving the appearance of a money pump. [If the state space had some representation of the set D, this paradox would also disappear.]*

In words, a person with the preferences in the previous problem would pay to move from a_2 to a_1, and then, conditional on each and every event in a partition of the state space, pay again to move back.

4.6.3. *Resolving the paradoxes.* In each of the problems above, the failure of countable additivity was to blame. One way to get around this failure is to put some flesh on the observation that "every finitely additive probability is the trace of a countably additive probability on a larger space." That is vague, but turns out to cover the essential idea behind one resolution of the paradoxes.

A bit of a warning here: This part touches on deep mathematics, the guidance that is given is close to a minimal logically necessary amount to do the one homework problem here. This can be uncomfortable, but try to see the structures of the arguments.

Fix a measure space $(X, \mathcal{X})$ (so that X is a non-empty set and $\mathcal{X}$ is a σ-field of subsets of X). There are deep Theorems (due to Stone) showing that there exists a compact Hausdorff[5] space $\widehat{X}$ and a mapping $\varphi : X \to \widehat{X}$ such that $\varphi(X)$ is dense in $\widehat{X}$, and, for each $E \in \mathcal{X}$, $\widehat{E}$, defined as the closure of $\varphi(E)$, is both a closed and an open subset of $\widehat{X}$. The space $\widehat{X}$ is called the **Stone space** for $(X, \mathcal{X})$.

Some useful facts about topological spaces (and the compact Hausdorff spaces are very useful topological spaces) for the next problem:

1. a set is open iff its complement is closed,
2. the finite union of closed sets is closed, equivalently, the finite intersection of open sets is open,
3. the empty set is both open and closed,
4. every closed subset of a compact space is compact, and finally,
5. if $(F_\alpha)_{\alpha \in A}$ is a collection of closed subsets of a compact space with $\cap_\alpha F_\alpha = \emptyset$, then $\cap_{\alpha' \in A'} F_{\alpha'} = \emptyset$ for some finite $A' \subset A$.

Homework 4.26. *Let $\widehat{\mathcal{X}}^\circ = \{\widehat{E} : E \in \mathcal{X}\}$, and let $\widehat{\mathcal{X}} = \sigma(\widehat{\mathcal{X}}^\circ)$. If P is a finitely additive probability on $\mathcal{X}$, define $\widehat{P}$ on $\widehat{\mathcal{X}}^\circ$ by $\widehat{P}(\widehat{E}) = P(E)$.*

1. *$\widehat{\mathcal{X}}^\circ$ is a field of subsets of $\widehat{X}$.*
2. *If P is a finitely additive probability on $\mathcal{X}$, then $\widehat{P}$ is a countably additive probability on $\widehat{\mathcal{X}}^\circ$, so has a unique countably additive extension to $\widehat{\mathcal{X}}$.*
3. *Suppose that $E_n \downarrow \emptyset$ in $\mathcal{X}$, but that $\lim_n P(E_n) > 0$. Show that $\cap_n \widehat{E}_n \neq \emptyset$ and that $\lim_n \widehat{P}(\widehat{E}_n) = \widehat{P}(\cap_n \widehat{E}_n)$. Compare this result with* Homework 3.39 *(if you took that detour).*
4. *An additional property of the Stone spaces is that for any csm (M, d), any measurable function $f : X \to M$, there exists a continuous function $\hat{f} : \widehat{X} \to M$ with the property that for any bounded, continuous $u : M \to \mathbb{R}$, $\int u(f(x))\, dP(x) = \int u(\hat{f}(\hat{x})\, d\widehat{P}(\hat{x})$.*

[5] A regularity condition that I am not going to explain here.

Let $\widehat{\mathbb{N}}$ be the Stone space for $(\mathbb{N}, 2^{\mathbb{N}})$ (which is isomorphic to the Stone-Čech compactification of the integers). For both of the money pumps given above, identify in $\widehat{\mathbb{N}}$ the location of the missing mass that makes the finitely additive money pumps possible.

5. Probabilities on Complete Separable Metric Spaces

Let (X, d) be a complete, separable metric (csm) space and $C_b(X)$ the set of bounded, continuous $\mathbb{R}$-valued functions on X. The supnorm metric on $C_b(X)$ is defined by

$$\rho(f, g) = \sup\{|f(x) - g(x)| : x \in X\}.$$

Lemma 5.1. *$(C_b(X), \rho)$ is a complete metric space. [(X, d) need not be complete or separable for this result.]*

Definition 5.2. *The space (X, d) has the* **finite intersection property** *if for every collection $\{F_\alpha : \alpha \in A\}$ of closed subsets of X with $\cap_{\alpha \in A} F_\alpha = \emptyset$, there is a finite $A' \subset A$ such that $\cap_{\alpha' \in A'} F_{\alpha'} = \emptyset$.*

Theorem 5.3 (FIP)**.** *(X, d) is compact iff it has the finite intersection property.*

Proof: Suppose that (X, d) has the fip and let x_n be a sequence in X. To show compactness, we must show that $\text{accum}(x_n) \neq \emptyset$. For each $n \in \mathbb{N}$, let $F_n = cl\,\{x_m : m \geq n\}$. For all finite $A' \subset \mathbb{N}$, $\cap_{n' \in A'} F_n \neq \emptyset$. Therefore, $\cap_n F_n \neq \emptyset$. But $\text{accum}(x_n) = \cap_n F_n$.

Suppose now that for any sequence x_n, $\text{accum}(x_n) \neq \emptyset$. Let $\{F_\alpha : \alpha \in A\}$ be a collection of closed subsets of X with $\cap_{\alpha \in A} F_\alpha = \emptyset$. For the purposes of establishing a contradiction, let us suppose that for all finite $B \subset A$, $\cap_{\beta \in B} F_\beta \neq \emptyset$.

Since $\cap_{\alpha \in A} F_\alpha = \emptyset$, we know that $\cup_{\alpha \in A} G_\alpha = X$ where $G_\alpha = F_\alpha^c$ is open. We need an intermediate step.

Lemma 5.4. *If (X, d) is separable, there is a countable collection, $\mathfrak{G} = \{G_n : n \in \mathbb{N}\}$, of open sets such that every open G is a countable union of the form $G = \cup_{n' \in \mathbb{N}'} G_{n'}$.*

Proof: Let X' be a countable dense subset of X and take $\mathfrak{G}$ to be the set $B(x', q)$, $x' \in X'$, $q \in \mathbb{Q}_{++}$. ∎

Back to the proof, from the Lemma, we know there exists a countable $A'' \subset A$ such that $\cup_{\alpha'' \in A''} G_{\alpha''} = X$. Therefore, $\cap_{\alpha'' \in A''} F_{\alpha''} = \emptyset$. Enumerate A'' as $(\alpha_k)_{k \in \mathbb{N}}$.

For each k, we know there exists an $x_k \in \cap_{m=1}^k F_{\alpha_m}$. Since each F_{α_m} is closed, $\text{accum}(x_k) \subset F_{\alpha_m}$. Therefore $\text{accum}(x_k) \subset \cap_m F_{\alpha_m}$. But $\text{accum}(x_k) \neq \emptyset$ contradicts $\cap_{\alpha'' \in A''} F_{\alpha''} = \emptyset$. ∎

5.1. **Some examples.**

5.1.1. $X = \mathbb{N}$. Let $X = \mathbb{N}$ and have the metric $e(x, y) = 0$ if $x = y$, $d(x, y) = 1$ if $x \neq y$.

Lemma 5.5. $2^{\mathbb{N}}$ *is uncountable.*

Proof: Any $E \in 2^{\mathbb{N}}$ can be identified with a point $s_E \in \{0, 1\}^\infty$ by defining $z_n(s_E) = 1_E(n)$, and any $s \in \{0, 1\}^\infty$ identifies an element $E_s \in 2^{\mathbb{N}}$ by $E_s = \{n \in \mathbb{N} : z_n(s) = 1\}$. We know that $\{0, 1\}^\infty$ is uncountable. ∎

Homework 5.1. $(\mathbb{N}, e)$ *is a csm,* $C_b(\mathbb{N})$ *consists of the set of all bounded functions on* $\mathbb{N}$*, and* $(C_b(\mathbb{N}), \rho)$ *is not separable. [For any* $E \in 2^{\mathbb{N}}$*,* $1_E(\cdot) \in C_b(\mathbb{N})$*, and if* $E \neq F$*, then* $\rho(1_E, 1_F) = 1$*.]*

5.1.2. $X = [0, 1]$. Let $X = [0, 1]$ and have the metric $d(x, y) = |x - y|$. Since X is compact, every continuous function on X is bounded so we omit the "$_b$" on $C(X)$.

Lemma 5.6. *If* $f \in C([0, 1])$*, then for every* $\epsilon > 0$ *there exists a* δ *such that for all* $x, y \in [0, 1]$*, if* $|x - y| < \delta$*, then* $|f(x) - f(y)| < \epsilon$*.*

Proof: Use the FIP Theorem. ∎

Homework 5.2. $(C([0, 1]), \rho)$ *is a csm.*

5.1.3. $X = \times_t \Omega_t$*, each* Ω_t *finite.* Let $\Omega = \times_{t \in \mathbb{N}} \Omega_t$ where each Ω_t is finite. For each t, define $\rho_t(\omega_t, \omega'_t)$ to be 1 if $\omega_t \neq \omega'_t$ and equal to 0 otherwise. Define a metric on Ω by

$$d(\omega, \omega') = \sum_t 2^{-t} \rho_t(z_t(\omega), z_t(\omega')).$$

Homework 5.3. *If* ω_n *is a sequence in* Ω*, then* $d(\omega_n, \omega) \to 0$ *iff for all* t*, there exists an* N *such that for all* $n \geq N$*,* $z_t(\omega_n) = z_t(\omega)$*. Further,* (Ω, d) *is compact.*

Let $\mathcal{C}^\circ$ be the field of cylinder sets in S^∞, S finite.

Homework 5.4. *Show that every cylinder set is closed. Using the finite intersection property, show that every finitely additive probability on $\mathcal{C}^\circ$ has a unique countably additive extension to $\mathcal{C} = \sigma(\mathcal{C}^\circ)$.*

Suppose now that each $\Omega_t = S$ for some finite S. Let $u : S \to \mathbb{R}$. For each $s \in \times_t S$ and $\beta \in (0,1)$, define $U_\beta(s) = \sum_t \beta^t u(z_t(s))$.

Homework 5.5. $U_\beta \in C(\times_t S)$.

If $(X_i, d_i)_{i \in I}$ is a finite collection of metric spaces, we define the product metric d on $X = \times_i X_i$ by $d(x, y) = \max_i d_i(x_i, y_i)$.

Homework 5.6. *If each $(X_i, d_i)_{i \in I}$ in a finite collection of metric spaces is compact, then so is (X, d).*

Consider a finite normal form game $\Gamma = (S_i, u_i)_{i \in I}$. Define $H^0 = \{h^0\}$ for some point h^0, and for $t \geq 1$, inductively define $H^t = \times_{\tau \leq t-1} S$. Let $\Sigma_{i,t}$ be the finite set $S_i^{H^t}$. Strategies for i in the infinitely repeated version of Γ are $\Sigma_i = \times_{t=0}^{\infty} \Sigma_{i,t}$. From Homework 5.3, we know that there is a nice metric d_i on Σ_i making (Σ_i, d_i) compact. From Homework 5.6, there is a metric d on $\Sigma = \times_i \Sigma_i$ making (Σ, d) compact. Let $\mathbb{O}(\sigma)$ be the outcome associated with play of the strategy vector $\sigma \in \Sigma$. Suppose that each $i \in I$ has a discount factor $0 < \beta_i < 1$. Define $U_i(\sigma) = \sum_t \beta_i^t u_i(z_t(\mathbb{O}(\sigma)))$.

Homework 5.7. $U_i(\cdot) \in C(\Sigma)$.

This means that infinitely repeated, finite games are a special case of compact metric space games.

Definition 5.7. *A game $\Gamma = (A_i, u_i)_{i \in I}$ is a compact metric space game if there exists metrics d_i such that*

1. *each (A_i, d_i) is a compact metric space, and*
2. *each $u_i \in C(A, d)$, $A = \times_i A_i$, $d(s, t) = \max_i d_i(s_i, t_i)$.*

5.2. **Borel probabilities.** With (X, d) a csm, let $\mathcal{X}$ be the σ-field generated by the open sets. A Borel probability is a countably additive probability on $\mathcal{X}$. The set of Borel probabilities on $(X, \mathcal{X})$ will be denoted $\Delta(X)$.

Recall that for $E \subset X$, $E^\epsilon = \cup_{x \in E} B(x, \epsilon)$ is the ϵ-ball around the set E. There are two, very different metrics on $\Delta(X)$. The variation norm (or strong) distance is

$$d_V(P, Q) = \sup_{E \in \mathcal{X}} |PE - QE|,$$

and the Prohorov (or weak) distance is

$$d_w(P, Q) = \inf\{\epsilon > 0 : (\forall E \in \mathcal{X})[PE < QE^\epsilon + \epsilon, \ \& \ QE < PE^\epsilon + \epsilon]\}.$$

Homework 5.8. *If $d_V(P^n, P) \to 0$, then $d_w(P^n, P) \to 0$. Let P^n be point mass on the point $1/n \in [0, 1]$ and let P be point mass on 0. Show that $d_w(P^n, P) \to 0$ but $d_V(P^n, P) \equiv 1$.*

It is a true fact (as opposed to that other kind of fact), that $d_w(P^n, P) \to 0$ iff $\int f \, dP^n \to \int f \, dP$ for all $f \in C_b(X)$.

Theorem 5.8. *If (X, d) is compact, then $(\Delta(X), d_w)$ is compact.*

Proof: Fill it in. ∎

5.3. **Consistency and learnability.** Suppose that (Θ, d) is a csm, and for each $\theta \in \Theta$ there is a distribution $\mu_\theta \in \Delta(X)$, $X \subset \mathbb{N}$. Let P_θ be the distribution on $X^{\mathbb{N}}$ given by i.i.d. draws from the distribution μ_θ. Let $Q \in \Delta(\Theta)$ be the prior distribution. Let Q_t be the Bayesian updating of Q after observing t draws from P_θ.

An interesting question is for what pairs (Q, μ_θ) does $d_w(Q_t, \delta_\theta) \to 0$ P_θ a.e. This is the question of the consistency of Bayes updating.

Another, closely related use of the word "consistency" shows up in statistics. Let $\widehat{\theta}_t \in \Theta$ be a sequence of estimators of θ, $\widehat{\theta}_t$ based on the first t observations from P_θ. The sequence of estimators is **consistent** if for all values of θ, $\widehat{\theta}_t \to \theta$ P_θ a.e.

In any case, consistency of Bayes updating in the JKR framework does not imply the learnability of μ_θ, and the learnability of μ_θ does not imply the consistency of Bayes updating.

5.4. **Compact metric space game.** Fix a compact metric space game $\Gamma = (A_i, u_i)_{i \in I}$. Let Δ_i be i's set of (Borel) mixed strategies, and let $\Delta = \times_i \Delta_i$. For each $\mu \in \Delta$, let $Br_i(\mu)$ denote i's set of mixed strategy best responses to μ, $Br_i^p(\mu)$ denote i's set of pure strategy best responses to μ.

Lemma 5.9. *For each $i \in I$, let X_i be a dense subset of A_i. $\mu \in \Delta = \times_i \Delta_i$ is an equilibrium iff for all $a_i \in X_i$, $u_i(\mu) \geq u_i(\mu \backslash a_i)$.*

Proof: Fill it in. ∎

Lemma 5.10. *Further, for each $\mu \in \Delta$, $Br_i^p(\mu)$ is a non-empty closed subset of A_i, and $Br_i(\mu)$ is the closed, convex set of probabilities putting mass 1 on $Br_i^p(\mu)$.*

Theorem 5.11. *Every compact metric space game has a non-empty, closed set of equilibria.*

Proof: First, non-emptiness.

Let $\epsilon_n \downarrow 0$. Let $X'_{i,n}$ be a finite ϵ_n-net for A_i. Let $X_{i,n} = \cup_{m \leq n} X'_{i,m}$ so that $X_{i,n}$ is also a finite ϵ_n-net for A_i. Let $X_i = \cup_n X_{i,n}$ so that for each i, X_i is dense in A_i.

Let $Eq(\Gamma_n)$ be the equilibrium set for the finite game $(X_{i,n}, u_i)_{i \in I}$. For each $n \in \mathbb{N}$, pick a $\mu_n \in Eq(\Gamma_n) \subset \Delta = \times_i \Delta_i(\S_i)$. Since Δ is compact, we know that $\text{accum}(\mu_n) \neq \emptyset$. Pick $\mu \in \text{accum}(\mu_n)$, and relabeling the sequence if necessary, assume that $d_w(\mu_n, \mu) \to 0$. We will show that μ is an equilibrium.

Suppose, for the purposes of establishing a contradiction, that μ is not an equilibrium. Then $\exists i \in I$, $\exists a_i \in X_i$, $\exists \epsilon > 0$ such that

$$u_i(\mu \backslash a_i) > u_i(\mu) + \epsilon.$$

We will show that for sufficiently large n, this implies that μ_n is not an equilibrium for Γ_n, establishing the contradiction.

We know that $u_i(\mu_n \backslash a_i) \to u_i(\mu \backslash a_i)$ and $u_i(\mu_n) \to u_i(\mu)$. Pick N_1 such that for all $n \geq N_1$, $|u_i(\mu_n \backslash a_i) - u_i(\mu \backslash a_i)| < \epsilon/3$ and $|u_i(\mu_n) \to u_i(\mu)| < \epsilon/3$. Note that this means that

$$u_i(\mu_n \backslash a_i) > u_i(\mu_n) + \epsilon/3.$$

Pick N_2 such that for all $n \geq N_2$, $a_i \in X_{i,n}$. For all $n \geq \max\{N_1, N_2\}$, μ_n is not an equilibrium by the last displayed inequality.

Second, closedness. Let μ_n be a sequence of equilibria converging to μ, if μ is not an equilibrium, repeat the previous logic with a couple of tiny changes. ∎

5.5. **Detour #4: Equilibrium Refinement for compact metric space games.**

5.5.1. *Perfect equilibria for finite games.* To begin with, let A be a finite set with the metric $d(a, b) = 1$ if $a \neq b$. Let $\mathcal{A}$ be the corresponding Borel σ-field. Note that (A, d) is compact, and that $\mathcal{A} = 2^A$.

Homework 5.9. *In this finite case, show that for* $\mu, \mu_n \in \Delta(\mathcal{A})$*,* $d_V(\mu_n, \mu) \to 0$ *iff* $d_w(\mu_n, \mu) \to 0$ *iff* $\sum_{a \in A} |\mu_n(a) - \mu_n(a)| \to 0$.

Let $\Gamma = (A_i, u_i)_{i \in I}$ be a finite game. For each $\Delta_i = \Delta(\mathcal{A}_i)$, define $d_i(\mu_i, \nu_i) = \sum_{a_i \in A_i} |\mu_i(a_i) - \nu_i(a_i)|$. For each $\mu \in \Delta = \times_i \Delta_i$, let $Br_i(\mu) \subset \Delta_i$ be the set of i's mixed best response to μ. Recall that $Br_i(\mu)$ is the convex hull of the pure strategy best responses to μ. Let $\Delta_i^{fs} \subset \Delta_i$ denote the set of **full support** μ_i, that is, the set of μ_i such that $\mu_i(a_i) > 0$ for each $a_i \in A_i$.

Definition 5.12 (Selten, Myerson)**.** *For* $\epsilon > 0$*, an* ϵ**-perfect equilibrium** *for a* Γ *is a vector* $\mu^\epsilon = (\mu_i^\epsilon)_{i \in I}$ *in* $\Delta^{fs} = \times_{i \in I} \Delta_i^{fs}$ *such that for each* $i \in I$*,*

$$d_i(\mu_i^\epsilon, Br_i(\mu^\epsilon)) < \epsilon. \tag{4}$$

A vector $\mu \in \Delta$ *is a* **perfect equilibrium** *if it is the limit as* $\epsilon_n \to 0$ *of* ϵ_n*-perfect equilibria.*

The requirement that each μ_i^ϵ be a full support distribution captures the notion that anything is possible, that any player may "tremble" and play any one of her actions. The requirement that each μ_i^ϵ be within d_i-distance ϵ of $Br_i(\mu^\epsilon)$ is, for finite games, equivalent to each agent i putting mass at least $1 - \epsilon$ on $Br_i(\mu^\epsilon)$. From Homework 5.9, as we send ϵ to 0, this is equivalent to both strong and weak closeness of the μ_i^ϵ to $Br_i(\mu^\epsilon)$. The situation is different for infinite games where the strong and the weak distances are very different, as you saw in Homework 5.8.

5.5.2. *Perfect equilibria for continuous payoff, compact metric space games.* Turning to infinite games, each A_i is assumed to be compact and each u_i is assumed to be jointly continuous on $\times_i A_i$. The set of mixed strategies for i, Δ_i, is the set of (Borel) probability measures on A_i, while Δ_i^{fs} is the set of probability measures assigning strictly positive mass to every non-empty open subset of A_i. Weak and strong distance from best response sets can be very different.

Homework 5.10. *Consider a single agent game played on* $[0, 1]$ *with continuous payoffs satisfying* $u(0) = 0$*,* $u'(x) = -1$ *for* $0 < x < \epsilon$ *and* $u'(x) = \frac{1}{2}\epsilon/(1 - \epsilon)$ *for* $\epsilon < x < 1$*.*

1. *Graph u (moderately carefully).*
2. *Show that point mass on 0 is the unique equilibrium strategy.*
3. *If* ν_i^ϵ *is the uniform distribution on the interval* $[0, \epsilon]$*, then* $d_w(\nu_i^\epsilon, Br_i) = \epsilon$ *but* $d_s(\nu_i^\epsilon, Br_i) = 1$.
4. *Show that* δ_ϵ*, point mass on* ϵ *is the worst choice, but satisfies* $d_w(\delta_\epsilon, Br_i) = \epsilon$*.*
5. *Characterize the set of* $\mu_i^\epsilon \in \Delta_i^{fs}$ *satisfying* $d_s(\mu_i^\epsilon, Br_i) < \epsilon$*.*

Definition 5.13. *A* **strong ϵ-perfect equilibrium** *is a vector $\mu^\epsilon = (\mu_i^\epsilon)_{i\in I}$ in Δ^{fs} such that for each $i \in I$,*

$$\rho_i^s(\mu_i^\epsilon, Br_i(\mu^\epsilon)) < \epsilon, \tag{5}$$

whereas a **weak ϵ-perfect equilibrium** *satisfies*

$$\rho_i^w(\mu_i^\epsilon, Br_i(\mu^\epsilon)) < \epsilon. \tag{6}$$

A vector $\mu \in \Delta$ is a **strong** *(respectively* **weak***)* **perfect equilibrium** *if it is the weak limit as $\epsilon^n \to 0$ of strong (respectively weak) ϵ^n-perfect equilibria.*

From Homework 5.9, strong and weak perfect equilibria are the same when the A_i are finite.

Let $\mathfrak{K}_Y$ denote the class of non-empty, compact subsets of a metric space (Y, d). For $A, B \in \mathfrak{K}_Y$, define $c(A, B) = \inf\{\epsilon > 0 : A \subset B^\epsilon\}$ where $B^\epsilon = \{y \in Y : \inf_{b\in B} d(y, b) < \epsilon\}$. The **Hausdorff distance** between compact sets is defined by

$$d_H(A, B) = \max\{c(A, B), c(B, A)\}.$$

Homework 5.11. *Suppose that (Y, d) is compact.*

1. *Every closed $F \subset Y$ belongs to $\mathfrak{K}_Y$.*
2. *Every finite subset of Y belongs to $\mathfrak{K}_Y$.*
3. *Show that every finite ϵ-net X^ϵ (see above) satisfies $d_H(X^\epsilon, Y) < \epsilon$.*
4. *The finite subsets of Y are d_H-dense in $\mathfrak{K}_Y$.*
5. *Show that $(\mathfrak{K}_Y, d_H)$ is a csm.*

Another way to define perfect equilibria for compact metric space games uses the limit-of-finite (lof) approximations approach. For $B_i \subset A_i$, $Br_i(B_i, \mu)$ denotes i's best responses to μ when i is constrained to play something in the set B_i.

Definition 5.14. *For each $i \in I$ and $\delta > 0$, B_i^δ denotes a finite subset of A_i within (Hausdorff distance) δ of A_i. For ϵ, a vector $\mu^{(\epsilon,\delta)} \in \times_{i\in I}\Delta_i^{fs}(B_i^\delta)$ is an* **(ϵ, δ)-perfect equilibrium with respect to** *$B^\delta = \times_{i\in I} B_i^\delta$ if for all $i \in I$,*

$$d_i^\delta(\mu_i^{(\epsilon,\delta)}, Br_i(B_i^\delta, \mu^{(\epsilon,\delta)})) < \epsilon, \tag{7}$$

where $d_i^\delta(\mu_i, \nu_i) = \sum_{a_i\in B_i^\delta} |\mu_i(a_i) - \nu_i(a_i)|$. We say that μ is a **limit-of-finite (lof) perfect equilibrium** *if it is the weak limit as $(\epsilon^n, \delta^n) \to (0, 0)$ of (ϵ^n, δ^n)-perfect equilibria with respect to some sequence B^{δ^n}.*

Homework 5.12. *Consider the 1 person game Γ with $A_i = \{0\}\times[0, 1]\cup\{1\}\times[0, 1] \subset \mathbb{R}^2$, and suppose that $u_i(x, r) = x$ for $x \in \{0, 1\}$, $r \in [0, 1]$. For each n, let $D_n = \{k/n : 0 \leq k \leq n\}$ and set $B_{i,n} = \{0\} \times D_{2^n} \cup \{1\} \times D_n$. For $p \in [1, \infty)$ and all finitely supported $\mu_i, \nu_i \in \Delta_i$, define the metrics $m_p(\mu_i, \nu_i) = (\sum_{a_i} |\mu_i(a_i) - \nu_i(a_i)|^p)^{1/p}$. Suppose that m_p is substituted for d_i^δ in Definition 5.14. For which values of p will every $(\epsilon_n, 1/n)$-equilibria converge to the equilibrium set of Γ?*

Definition 5.15. *A pure strategy, $a_i \in A_i$ is* **weakly dominated** *for i if there exists a mixed strategy, $\mu_i \in \Delta_i$ such that for all $a \in A$, $u_i(a\backslash a_i) \leq u_i(a\backslash \mu_i)$ and for some $a' \in A$, $u_i(a'\backslash a_i) < u_i(a'\backslash \mu_i)$. A vector $\mu \in \Delta$ is* **limit admissible** *if for all $i \in I$, $\mu_i(\mathcal{O}_i) = 0$, where $\mathcal{O}_i$ denotes the interior of the set of strategies weakly dominated for i.*

The following problems are stylized versions of a differentiated commodity Bertrand pricing game in which agent i's best response is always to undercut agent j by a finite amount. Players' payoffs in these examples are based on the following continuous function on $[0, \frac{1}{2}] \times [0, \frac{1}{2}]$.

$$v(x, y) = \begin{cases} x & \text{if } x \leq \frac{1}{2}y \\ \frac{y(1-x)}{2-y} & \text{if } \frac{1}{2}y < x \end{cases} \tag{8}$$

You should graph a couple of sections of this function to see what is going on. We will think of x as agent i's strategy and y as agent j's strategy. Note that for all x and y, $v(x, y) \geq 0$, and if either $x = 0$ or $y = 0$, then $v(x, y) = 0$. Thus, i is indifferent between all actions when $y = 0$. If $y > 0$, then $v(\cdot, y)$ increases from 0 with slope 1 to its unique maximum at $x = \frac{1}{2}y$, and decreases linearly on $(\frac{1}{2}y, \frac{1}{2}]$. (The negative slope is chosen so that $v(1, y) = 0$.) Thus, for $y > 0$, the unique solution to the problem $\max\{v(x, y) : x \in [0, \frac{1}{2}]\}$ is $x = \frac{1}{2}y$.

Homework 5.13. *$A_1 = A_2 = [0, \frac{1}{2}]$, and the utility functions are given by $u_i(a_i, a_j) = v(a_i, a_j)$ where v is given above.*[6] *Show that the unique equilibrium for this game is $(a_1, a_2) = (0, 0)$, but for each agent, the strategy $a_i = 0$ is weakly dominated.*

This shows that putting mass 0 on weakly dominated strategies and equilibrium existence are not compatible.

Homework 5.14. *Let $A_1 = A_2 = [-\frac{1}{2}, \frac{1}{2}]$. Set $u_1(a_1, a_2) = u_2(a_1, a_2) = 0$ if either a_1 or a_2 is in $[-\frac{1}{2}, 0)$, otherwise let the payoffs be as in Homework 5.13. Show that*

1. *The strategy $\mu = (\mu_1, \mu_2)$ is a Nash equilibrium if $\mu_i([-\frac{1}{2}, 0]) = 1$, $i = 1, 2$.*
2. *The interior of i's weakly dominated strategies is $[-\frac{1}{2}, 0)$, so any refinement of Nash equilibrium that satisfies existence and is limit admissible puts mass 1 on the point $(0, 0)$.*
3. *All of the weakly dominated strategies are equivalent.*

It is clear that every strong perfect equilibrium is a weak perfect equilibrium because $d_w(\mu, \nu) \leq d_s(\mu, \nu)$. The inclusion can be strict.

Homework 5.15. *Consider the two person game Γ with $A_1 = \{-1\} \cup [0, 1]$ and $A_2 = [0, 1]$. Agent 2's payoffs are strictly decreasing in her own actions and independent of 1's actions: $u_2(a_1, a_2) = -a_2$, while Agent 1's payoffs are is given by*[7]

$$u_1(a_1, a_2) = \begin{cases} \frac{1}{8}a_2 & \text{if } a_1 = -1 \\ a_1 & \text{if } a_1 \in [0, \frac{1}{2}a_2) \\ a_2 - a_1 & \text{if } a_1 \in [\frac{1}{2}a_2, 1] \end{cases}$$

(In a continuous time entry game interpretation of this model, $a_1 = -1$ corresponds to the first firm entering the market long before the second firm can.)

This problem asks you to fill in the steps to prove:

In any Nash equilibrium for Γ, 2 puts mass 1 on her strict best response set, $\{0\}$, and 1 puts mass 1 on the two point set $\{-1, 0\}$. The only strong perfect equilibrium for this game is $(a_1, a_2) = (-1, 0)$, while $(0, 0)$ is a weak perfect equilibrium.

1. *Verify that the Nash equilibrium set is as described.*
2. *$(-1, 0)$ is the unique strong perfect equilibrium: let $(\mu_1^\epsilon, \mu_2^\epsilon)$ be a strong ϵ-perfect equilibrium. Because 0 is 2's strict best response, $\mu_2^\epsilon(\{0\}) \geq 1 - \epsilon$. Show that, for small ϵ, 1's payoff to any $a_1 \geq 0$ is less than or equal to 0 against any such μ_2^ϵ. By contrast, show that against any such μ_2^ϵ, 1's payoffs to $a_1 = -1$ is strictly positive. Taking limits, show that $(-1, 0)$ is the unique strong perfect equilibrium.*
3. *$(0, 0)$ is a weak perfect equilibrium: show that it is possible to construct full support distributions for agent 2 that have two properties: they put mass greater than or equal to $1 - \epsilon$ on a 2ϵ-neighborhood of 2's strict best response set; and 1's best response is strictly positive. [Pick $\epsilon > 0$. Let ν_2^ϵ denote a full support distribution and set $\mu_2^\epsilon = (1 - \epsilon) \cdot \delta_\epsilon + \epsilon \cdot \nu_2^\epsilon$ where δ_ϵ denotes point mass on the point ϵ in A_2. Against μ_2^ϵ, agent 1's payoff to $a_1 = -1$ is equal to $\frac{1}{8}$ times the mean of μ_2^ϵ, and this is bounded above by $\frac{1}{8}[(1 - \epsilon) \cdot \epsilon + \epsilon \cdot 1] = \frac{1}{8}[2 \cdot \epsilon - \epsilon^2] = \frac{1}{4}\epsilon + o$, where o is a second order term in ϵ. To calculate a lower bound for agent 1's payoff to playing $a_1 = \frac{1}{2}\epsilon$ against μ_2^ϵ, note that $u_1(\frac{1}{2}\epsilon, \cdot) \geq -\frac{1}{2}\epsilon$. Thus, agent 1's payoff to $a_1 = \frac{1}{2}\epsilon$ is greater than or equal to $(1 - \epsilon)\frac{1}{2}\epsilon + \epsilon(-\frac{1}{2}\epsilon) = \frac{1}{2}\epsilon - o$, strictly greater than $\frac{1}{4}\epsilon + o$ for small ϵ, so 1's best response is strictly positive.]*

Homework 5.16. *$A_1 = A_2 = \{-1\} \cup [0, 1]$. The utility functions are symmetric,*

$$u_i(a_i, a_j) = \begin{cases} 0 & \text{if } a_i = -1 \\ 2 & \text{if } a_i, a_j \in [0, 1] \\ -a_i & \text{if } a_j = -1 \text{ and } a_i \in [0, 1] \end{cases}$$

1. *The strategy $a_i = 0$ weakly dominates every other strategy.*
2. *$(a_1, a_2) = (-1, -1)$ is a* lof *perfect equilibrium. [For $i = 1, 2$, let B_i^n be a sequence of a finite approximations converging to A_i such that for all $n \in \mathbb{N}$, $(0, 0) \notin (B_1^n, B_2^n)$. If j is playing $a_j = -1$, then because B_i^n does not contain the point 0, $a_i = -1$ is a strict best response.]*
3. *Verify that $\{-1\}$ is an open subset of the set of weakly dominated strategies. This means that this* lof *perfect equilibrium violates limit admissibility.*

Definition 5.16. *For $i \in I$, let F_i denote a finite subset of A_i and let F denote $\times_{i \in I} F_i$.*

(*a*) *The sequence of approximations B^n is* **anchored at** *F if $F \subseteq B^n$ for all $n \in \mathbb{N}$.*

(*b*) *A vector of strategies $\mu = (\mu_i)_{i \in I}$ is a* **lof perfect equilibrium anchored at** *F if it satisfies Definition 5.14 above, with the added restriction that the sequence of approximations, B^{δ^n}, be anchored at F.*

(c) *A vector of strategies* $\mu = (\mu_i)_{i \in I}$ *is an* **anchored perfect equilibrium** *if* $\mu \in \cap_F Per(F)$ *where* $Per(F)$ *denotes the set of lof perfect equilibria anchored at* F *and the intersection is taken over all finite* F.

Anchored perfect equilibria are immune to the inclusion of any finite set of pure strategies in the sequence of finite approximations to the infinite strategy spaces.

Homework 5.17. *Show that* $(-1, -1)$ *is* **not** *an anchored lof perfect equilibrium in Homework* 5.16.

There is improvement in anchoring the lof approach, but it still does not rid us of many weakly dominated equilibria.

Homework 5.18. *In this two firm entry game,* (γ, t) *represents entry in market* γ *at time* t, $\gamma = \alpha, \beta$. *Firms have resources sufficient to enter only one market. Firm* 2 *is indifferent between markets and times of entry, while firm* 1 *wishes to enter market* α *if* 2 *enters, and wishes to enter market* β *at the same time as* 2 *if* 2 *enters that market. The pure strategies are* $A_1 = A_2 = \{\{\alpha\} \times [0,1]\} \cup \{\{\beta\} \times [0,1]\}$ *with typical element* (m_i, a_i), $m_i \in \{\alpha, \beta\}$, $a_i \in [0,1]$. 2*'s utility function is constant at* 0. 1*'s utility function is* $u_1((m_1, a_1), (\beta, a_2)) = -|a_1 - a_2|$, *while* $u_1((\alpha, a_1), (\alpha, a_2)) = 0$ *and* $u_1((\beta, a_1), (\alpha, a_2)) = -1$.

1. *For every* $a_1 \in [0,1]$, *the strategy* (α, a_1) *is weakly dominated by* (β, a_1) *and by no other strategy.*
2. *No* (β, a_1) *is weakly dominated for* 1.
3. $((\alpha, a), (\alpha, a))$ *is an anchored perfect equilibrium for any* $a \in [0,1]$. [*Fix an arbitrary* $a \in [0,1]$ *and finite set* $F = F_1 \times F_2 \subset A_1 \times A_2$. *Let* $S \subset [0,1]$ *be the set of points,* s, *such that* $(m_i, s) \in F_i$ *for some* i *and/or some* m_i. *Pick two sequences of finite subsets of* $[0,1]$, C^n *and* D^n *converging to* $[0,1]$, *such that* C^n, D^n *and* S *are pairwise disjoint. Let* $B_i^n = \{\{\alpha\} \times C^n\} \cup \{\{\beta\} \times D^n\}$ *for* $i = 1, 2$. *Choose* c^n *in* C^n *converging to* a. *Because* (α, c^n) *is a strict best response for 1 against the play of* (α, c^n) *by 2,* $((\alpha, c^n), (\alpha, c^n))$ *is a perfect equilibrium for the finite game played on* $B_1^n \times B_2^n$.]

5.5.3. *Proper equilibria for finite games.* From the musty recesses of your brain, pull out the following

Definition 5.17 (Myerson)**.** *For a finite game,* $\mu^\epsilon \in \Delta$ *is an* ϵ*-proper equilibrium if*

(a) *it is an* ϵ*-perfect equilibrium, and*
(b) *for all* $i \in I$, $a_i, b_i \in A_i$, *if* $u_i(\mu^\epsilon \backslash a_i) < u_i(\mu^\epsilon \backslash b_i)$, *then* $\mu_i^\epsilon(a_i) \leq \epsilon \cdot \mu_i^\epsilon(b_i)$.

A vector $\mu \in \Delta$ *is a proper equilibrium if it is the limit as* $\epsilon^n \to 0$ *of* ϵ^n*-proper equilibria.*

Enough of that finite stuff.

5.5.4. *LOF proper equilibria for continuous payoff, compact metric space games.* From the *lof* perspective, there is no problem defining properness: we simply replace the word "perfect" in Definition 5.14 with "proper." For finite games, proper equilibria are a non-empty subset of the perfect equilibria, so the same holds for lof proper equilibria or anchored lof proper equilibria. For *lof* proper equilibria, the choice of a particular large finite game may determine the set of predictions, even in the anchored approach.

Homework 5.19. $A_1 = A_2 = [-1, +1]$. *1's payoffs achieve a strict maximum at* $a_1 = 0$, $u_1(a_1) = -|a_1|$. *2's payoffs are given by* $u_2(a_1, a_2) = a_1 \cdot a_2$.

1. *The Nash equilibria for the game involve* 1 *playing* 0 *and* 2 *playing any mixed strategy.*
2. *For every anchoring set* F, *there is a sequence* $B^n \supseteq F$ *of finite approximations to* A *such that* $(0, +1)$ *is the only limit of proper equilibria for the games played on* B^n.
3. *For every anchoring set* F, *there is a sequence* $B^n \supseteq F$ *of finite approximations to* A *such that* $(0, -1)$ *is the only limit of proper equilibria for the games played on* B^n.

5.5.5. *Weak and strong proper equilibria for continuous payoff, compact metric space games.* It may not be possible to simultaneously satisfy infinitely many relative weight conditions on a mixed strategy.

Example 5.1. *There is a single agent whose action space is* $[0, 2]$. *Her strictly decreasing utility function is* $u(a) = -a$ *so that the unique Nash equilibrium is* 0. *For the partition* $\mathcal{A} = \{[0, \frac{1}{2}), [\frac{1}{2}, \frac{3}{4}), \ldots, [1, 1\frac{1}{2}), [1\frac{1}{2}, 1\frac{3}{4}), \ldots, \{2\}\}$ *of* $[0, 2]$, *there is no* $\epsilon \in (0, 1)$ *and full support distribution,* μ, *on* $[0, 2]$ *with the property that* $\mu(A) \leq \epsilon \cdot \mu(B)$ *for all pairs* $A, B \in \mathcal{A}$ *with* $u(A) \ll u(B)$ *(where for* $S, T \subset \mathbb{R}$, *we write* $S \ll T$ *if the supremum of the numbers in* S *is less than the infimum of the numbers in* T*).*

The resolution of this difficulty is to require that the relative weight conditions hold for finite measurable partitions of the action spaces. The final part of the definition requires that the set of proper equilibria not depend on any particular finite partition by 'anchoring' the finite partitions.

Definition 5.18. *Let* $\epsilon > 0$ *and* $\mathcal{P} = (\mathcal{P}_i)_{i \in I}$ *denote a vector of finite partitions of* $(A_i)_{i \in I}$. *We say that a vector of strategies,* $\mu = \mu^\epsilon(\mathcal{P})$, *is a* **strong** *(***weak***)* **ϵ-proper equilibrium relative to** $\mathcal{P}$ *if it is*

(*a*) *a strong (weak)* ϵ*-perfect equilibrium, and if*
(*b*) *for all* $i \in I$, *if* $u_i(\mu \backslash R_i) \ll u_i(\mu \backslash S_i)$, $R_i, S_i \in \mathcal{P}_i$, *then* $\mu_i(R_i) \leq \epsilon \cdot \mu_i(S_i)$.

We say that μ *is a* **strong** *(***weak***)* **proper equilibrium relative to** $\mathcal{P}$ *if it is the limit of strong (weak)* ϵ^n*-proper equilibria relative to* $\mathcal{P}$, $\epsilon^n \to 0$. *Finally, a vector of strategies,* $\mu = (\mu_i)_{i \in I}$, *is a* **strong** *(***weak***)* **proper equilibrium** *if* $\mu \in \cap_{\mathcal{P}} Pro^s(\mathcal{P})$ *(*$\mu \in \cap_{\mathcal{P}} Pro^w(\mathcal{P})$*) where* $Pro^s(\mathcal{P})$ *(*$Pro^w(\mathcal{P})$*) denotes the strong (weak) proper equilibria relative to* $\mathcal{P}$ *and the intersection is taken over all finite measurable partitions* $\mathcal{P}$.

There are equilibria that are weakly proper even though they are not even strongly perfect.

Homework 5.20. *Show that the strategies* $(0,0)$ *are a weak* proper *equilibrium in Homework* 5.15. [*Fix a measurable partition* $\mathcal{P}_2 = \{P_{2,1}, \dots, P_{2,k}\}$ *of* $A_2 = [0,1]$. *The strategy of the proof is to take a sequence of normal random variables with mean* ϵ *and variance* ϵ^2*, condition their densities to the interval [0, 1], and to perturb the resulting random variable so that each element of* $\mathcal{P}_2$ *is assigned positive mass. Choose the perturbation so that as* ϵ *converges to* 0 *the relative probability relations required by properness are satisfied. In response to a distribution which is nearly point mass at* ϵ*, the payoffs to agent* 1 *of playing* -1 *are essentially* $\frac{1}{8}\epsilon$*, while the payoffs to playing* $\frac{1}{2}\epsilon$ *are essentially* $\frac{1}{2}\epsilon$ *so that* 1*'s best response set is strictly positive.*]

5.5.6. *One of the existence and closure proofs.* The following shows of one more use of the fip property characterization of compactness.

Theorem 5.19. *The set of anchored perfect (proper) equilibria is a closed, nonempty set of the Nash equilibria.*

Homework 5.21. *Using the following outline, prove Theorem 5.19.*

1. *For* $\epsilon, \delta > 0$*, let* $\text{cl}\, P(\epsilon, \delta, F)$ *denote the closure of the set of* ϵ*-perfect (resp. proper) equilibria for finite games where each* $i \in I$ *uses the strategy set* $B_i^\delta \supseteq F_i$ *within Hausdorff distance* δ *of* A_i*. By Selten [1975] (resp. Meyerson [1978]), this set is not empty.*

   Show that the collection $\{\text{cl}\, P(\epsilon, \delta, F) : \epsilon > 0,\ \delta > 0\}$ *has the finite intersection property.*
2. *Because* Δ *is compact, the set* $P(F) := \bigcap_{\epsilon,\delta>0}\{\text{cl}\, P(\epsilon, \delta, F)\}$ *is not empty. To finish the proof for perfect (proper) equilibria anchored at* F*, show that*
   (a) $P(F)$ *is a subset of the Nash equilibria,*
   (b) $P(F)$ *is equal to the set of perfect (resp. proper) equilibria anchored at* F*.*
3. *Show that the collection* $\{P(F) : F \text{ a finite subset of } A\}$ *has the finite intersection property in the compact set* Δ*. Hence the set of anchored perfect equilibria,* $\bigcap_F P(F)$*, is not empty.*

5.5.7. *Questions about infinitely repeated finite games.* Let $(S_i, u_i)_{i\in I}$ be a finite game and $\mu = (\mu_i)_{i\in I}$ a proper equilibrium for $(S_i, u_i)_{i\in I}$. Let Γ be the compact metric space game with continuous payoffs that arise when $(S_i, u_i)_{i\in I}$ is repeated infinitely often and payoffs to the history $h \in S^\infty$ are given by $U_i(h) = \sum_t (\beta_i)^t u_i(z_t(h))$, $0 < \beta_i < 1$.

Question: What do the finite ϵ-nets of the repeated game strategy sets look like? [This is known, see [8].]

Question: is $\sigma_{i,t} \equiv \mu_i$ a strong (weak, lof, anchored lof) proper equilibrium? [This is not known so far as I know, but I'd bet the answer is yes in each case except, possibly, the lof proper case.]

5.5.8. *Stability by Hillas.* A gtc (game theory correspondence) from a compact, convex metric space to itself is one that is non-empty valued, convex valued, and has a closed graph. Such correspondences are known to have fixed points. From this one can derive the existence of Nash equilibria in compact metric space games just as one does for finite games.

Define the strong Hillas distance between two gtc's mapping Δ to Δ by

$$\rho_s(\Psi, \Psi') = \sup_{x \in X} d_{H,s}(\Psi(x), \Psi'(x)),$$

where $d_{H,s}$ is the Hausdorff distance using d_s to measure the distance between strategies. To define the weak Hillas distance between two gtc's, replace $d_{H,s}$ by $d_{H,w}$,

$$\rho_w(\Psi, \Psi') = \sup_{x \in X} d_{H,w}(\Psi(x), \Psi'(x)),$$

where $d_{H,w}$ is the Hausdorff distance using d_w to measure the distance between strategies.

Let Br be the correspondence $\mu \mapsto \times_i Br_i(\mu)$.

Homework 5.22. *Br is a gtc.*

Definition 5.20. *A closed set $E \subset Eq(\Gamma)$ has the strong (respectively weak) property* (S) *if it satisfies*

(S) *for all sequences of gtc's Ψ^n, $\rho_s(\Psi^n, Br) \to 0$, (respectively $\rho_w(\Psi^n, Br) \to 0$), there exists a sequence σ^n of fixed points of Ψ^n such that $d_w(\sigma^n, E) \to 0$.*

A closed set $E \subset Eq(\Gamma)$ is strongly (respectively weakly) Hillas stable if it has the strong (respectively weak) property (S) *and no closed, non-empty, proper subset of E has the strong (respectively weak) property* (S).

This can be said as "E is (Hillas) stable if it is minimal with respect to the strong (weak) property (S)." It can be shown that

Theorem 5.21. *Strong (weak) Hillas stable sets exist for compact, continuous games. Further, every strong (weak) Hillas stable set is a subset of the strongly (weakly) perfect equilibria and contains a strongly (weakly) proper equilibrium.*

However, the only hard copy of the proof is lost, and electronic copies cannot be found either.

5.6. **Detour #5: Stochastic versions of Berge's Theorem of the Maximum.** Fix a probability space $(\Omega, \mathcal{F}, P)$. Let (Θ, d) be a compact metric space and $C(\Theta)$ the set of continuous, real-valued functions on Θ. Let $\mathcal{C}$ denote the Borel σ-field on $C(\Theta)$.

For $f, g \in C(\Theta)$, and $\alpha, \beta \in \mathbb{R}$, we define the functions $\alpha f + \beta g$ and $f \cdot g$ by

$$(\alpha f + \beta g)(x) = \alpha f(x) + \beta g(x), \; (f \cdot g)(x) = f(x) \cdot g(x).$$

Homework 5.23. *If $f, g \in C(\Theta)$, then $\alpha f + \beta g, f \cdot g \in C(\Theta)$.*

Definition 5.22. *A class of functions $\mathcal{A} \subset C(\Theta)$ is an* **algebra** *if for $f, g \in \mathcal{A}$, and $\alpha, \beta \in \mathbb{R}$, the functions $\alpha f + \beta g, f \cdot g \in \mathcal{A}$. The class $\mathcal{A}$* **separates points** *if for all $\theta \neq \theta'$, there is a functon $f \in \mathcal{A}$ such that $f(\theta) \neq f(\theta')$. The class $\mathcal{A}$* **contains the constant functions** *if for all $\alpha \in \mathbb{R}$, $\alpha \cdot \mathbf{1} \in \mathcal{A}$ where $\mathbf{1}$ is the function identically equal to 1.*

Remember that $C(\Theta)$ has the metric ρ defined by $\rho(f,g) = \max_\theta |f(\theta) - g(\theta)|$. We can substitute "max" for "sup" because we've assumed that Θ is compact.

The following is very important. We'll use it for some relatively trivial stuff, but we won't prove it.

Theorem 5.23 (Stone-Weierstrass)**.** *If Θ is compact and $\mathcal{A} \subset C(\Theta)$ is a dense subset of an algebra that separates points and contains the constants, then $\text{cl}\,\mathcal{A} = C(\Theta)$.*

The following uses the Stone-Weierstrass theorem to show that $(C(\Theta), \rho)$ is a csm when Θ is compact.

Homework 5.24. *Let Θ' be a countable dense subset of Θ. For each $\theta' \in \Theta'$ and each rational $q \geq 0$, define $f_{\theta',q}(\theta) = \max\{1 - qd(\theta,\theta'), 0\}$.*

1. *Show that $f_{\theta',q} \in C(\Theta)$.*
2. *Show that the collection $\mathcal{A}' = \{f_(\theta', q) : \theta' \in \Theta',\ q \in \mathbb{Q}_+\}$ separates points and contains the constants.*
3. *Let $\mathcal{P}_{n,\mathbb{Q}}$ denote the set of polynomials of degree n having rational coefficients. For all n, if $p \in \mathcal{P}_{n,\mathbb{Q}}$ and $f_1, \ldots, f_n \in C(\Theta)$, then $p(f_1, \ldots, f_n) \in C(\Theta)$.*
4. *Show that $\cup_n \mathcal{P}_{n,\mathbb{Q}}(\mathcal{A}')$ is a countable set that is dense in an algebra that separates points and contains the constants.*
5. *$(C(\Theta), \rho)$ is a csm.*

Definition 5.24. *The* **evaluation mapping** *is the function $e : C(\Theta) \times \Theta \to \mathbb{R}$ defined by*

$$e(f, \theta) = f(\theta).$$

Remember that product spaces are given product metrics, in particular, $C(\Theta) \times \Theta$ is given the metric $d((f,\theta),(g,\theta')) = \max\{\rho(f,g), d(\theta,\theta')\}$

Homework 5.25. *The evaluation mapping is continuous.*

Let $X : \Omega \to C(\Theta)$ be a random variable, that is, for all $E \in \mathcal{C}$, $X^{-1}(E) \in \mathcal{F}$. For $\omega \in \Omega$, let X_ω be the value of X at ω. We are interested in the stochastic maximization problem

$$\max_{\theta \in \Theta} X_\omega(\theta),$$

and the behavior of the related

$$\Psi(\omega) := \{\theta^* \in \Theta : (\forall \theta' \in \Theta)[X_\omega(\theta^*) \geq X_\omega(\theta)]\}.$$

For $f \in C(\Theta)$,

$$\Psi(f) := \{\theta^* \in \Theta : (\forall \theta' \in \Theta)[f(\theta^*) \geq f(\theta)]\}.$$

Thus, we are using $\Psi(\omega)$ as short-hand for $\Psi(X_\omega)$.

Homework 5.26. *Suppose that $\Psi(f)$ contains only one element, call it θ_f. For every $\epsilon > 0$, there exists a $\delta > 0$ such that for all g satisfying $\rho(f,g) < \delta$, $d(\theta_f, \Psi(g)) < \epsilon$.*

Theorem 5.25. *If $X_n : \Omega \to C(\Theta)$ is a sequence of random variables, $P(X_n \to f) = 1$, $\hat{\theta}_n(\omega)$ is a measurable function with the property that $P(\hat{\theta}_n \in \Psi(X_n(\omega)) = 1$, and $\Psi(f)$ contains only one element, call it θ_f, then $P(htheta_n \to \theta_f) = 1$.*

Homework 5.27. *Prove Theorem 5.25.*

Homework 5.28. *Let $E_S \subset C(\Theta)$ denote the set of f such that $\Psi(f)$ contains only one element.*

1. *Show that $E_S \in \mathcal{C}$.*
2. *Show that the function $\hat{\theta} : E_S \to \Theta$ defined by $\{\hat{\theta}(f)\} = \Psi(f)$ is continuous, hence measurable.*

Theorem 5.26. *Suppose that $X : \Omega \to C(\Theta)$ satisfies $P(X \in E_S) = 1$. If $X_n : \Omega \to C(\Theta)$ is a sequence of random variables, $P(X_n \to X) = 1$, $\hat{\theta}_n(\omega)$ is a measurable function with the property that $P(\hat{\theta}_n \in \Psi(X_n(\omega)) = 1$, then $P(\hat{\theta}_n \to \Psi(X)) = 1$.*

Homework 5.29. *Prove Theorem 5.26.*

[THIS DETOUR IS NOT QUITE FINISHED YET]

6. Fictitious Play and Related Dynamics

Fictitious play gives a deterministic dynamic process with a state space which is the product of an infinite and a finite state space. We are mostly, but not exclusively, interested in the behavior of the finite part of the state space. For these purposes, fix a finite game $\Gamma = (S_i, u_i)_{i \in I}$ and let $S = \times_i S_i$.

6.1. The basics.

The "beliefs" of each $i \in I$ at times $t \in \{0, 1, 2, \dots\}$ are points $\gamma_t^i \in \Delta^{fs}(S_{-i})$ where for any finite set E, $\Delta^{fs}(E) = \{m \in \mathbb{R}_{++}^E : \sum_{e \in E} m(e) = 1\}$ is the set of strictly positive probabilities on E. The "weight" given to beliefs by i at time t is $w_t^i \in \mathbb{R}_{++}$. Given beliefs $\gamma_t = (\gamma_t^i)_{i \in I}$, a vector $s_t \in \times_i Br_i^P(\gamma_t^i)$ is picked. To be complete, if more than one of i's pure strategies are indifferent given beliefs γ_t^i, i will pick according to some ordering of the points in S_i. We now specify how the vector (γ_t, w_t) is updated. If at time t, the vector s happens, then i's beliefs-weight vector at time $t+1$ is

$$(\gamma_{t+1}^i, w_{t+1}^i) = (\frac{w_t^i}{w_t^i + 1}\gamma_t^i + \frac{1}{w_t^i + 1}\delta_{s_{-i}}, w_t^i + 1).$$

Let $w_t = (w_t^i)_{i \in I}$. The whole dynamic process (s_t, γ_t, w_t) is specified once the initial conditions (γ_0, w_0) are given. This class of dynamic processes is called "fictitious play."

Letting $e_{s_{-i}} \in \mathbb{R}^{S_{-i}}$ denote the unit vector in the s_{-i} direction. Setting $\kappa_t^i(s_{-i}) = w_t^i \gamma_t^i(s_{-i})$ and $\kappa_{t+1}^i = \kappa_t^i + e_{s_{-i}}$ gives another formulation of the dynamic that is sometimes easier to keep track of since one simply adds 1 to $\kappa_t^i(s_{-i})$ if s_{-i} happens, and add 0 otherwise.

Definition 6.1. *A pure strategy equilibrium $s^* \in S$ is* **strict** *if for all $i \in I$, $Br_i^P(s^*) = \{s_i^*\}$.*

Homework 6.1. *Suppose that s^* is a strict equilibrium for Γ. Show that for each $i \in I$, there is an open $G_{-i} \subset \Delta^{fs}(S_{-i})$ containing $\delta_{s^*_{-i}}$ such that if there exists a T with $\gamma_T \in \times_i G_{-i}$, then for all $t \geq T$, $\gamma_t \in \times_i G_{-i}$ and $s_t = s^*$.*

Given any sequence $s \in S^\infty$, we construct the sequence D_t of empirical distributions as follows:

$$D_t(a) = \tfrac{1}{t} \sum_{\tau=1}^{t} 1_{z_\tau(s)=a},$$

so that $D_t \in \Delta(S)$. For each $i \in I$ and $D_t \in \Delta(S)$, define $D_t^i \in \Delta(S_{-i})$ to be the marginal distribution of D_t on S_{-i}, that is, by

$$D_t^i(s_{-i}) = \sum_{t_i \in S_i} D_t(t_i, s_{-i}).$$

The following problem should be compared with Homework 6.1

Homework 6.2. *If $D_t \to \delta_{s^*}$ and s results from fictitious play, then s^* is an equilibrium of Γ.*

If s is arbitrary, in particular, if it need not come from fictitious play, then the behavior of the sequence D_t in the compact metric space $\Delta(S)^\infty$ can be pretty arbitrary.

Homework 6.3. *Without assuming that s results from fictitious play, give an $s \in S^\infty$*

1. *such that s is not convergent but D_t converges to a point in $\Delta(S)$,*
2. *such that D_t is non-convergent,*
3. *such that $\text{accum}(D_t) = \Delta(S)$, and*
4. *such that D_t is non-convergent, but $Q_t := \frac{1}{t}\sum_{\tau=1}^{t} D_t$ is convergent.*

Homework 6.4. *If $s \in S^\infty$ results from fictitious play starting at arbitrary initial conditions (γ_0, w_0), then for all $i \in I$, i's beliefs are* **asymptotically empirical**, *that is, $\|\gamma_t^i - D_t^i\| \to 0$. [Note that this is true whether or not D_t converges.]*

Homework 6.5. *Consider the 2×2 game*

	Left	Right
Up	$(0,0)$	$(1,1)$
Down	$(1,1)$	$(0,0)$

Find the sets of initial conditions (γ_0, w_0) for which the corresponding fictitious play process has the property

1. *that each D_t^i converges,*
2. *that D_t converges, and*
3. *that D_t converges to a Nash equilibrium.*

For $\nu \in \Delta(S)$, let $\text{marg}_{S_i}(\nu)$ be the marginal distribution of ν on S_i.

Lemma 6.2. *If for all $i \in I$, $\text{marg}_{S_i}(D_t) \to \sigma_i$, then $(\sigma_i)_{i \in I}$ is a Nash equilibrium.*

6.2. **Bayesian updating and fictitious play.** One of the interpretations of fictitious play is that all the players are convinced that everyone else is playing some iid mixed strategy. We know, from Nachbar [18], that optimization against correct beliefs is difficult to arrange unless one starts with an equilibrium. Here, we've got a model of players who act a bit psychotically — they believe that everyone else is an automaton, and may persist in this belief in the face of a huge amount of evidence to the contrary. Before going through that interpretation in detail, it is worth "reviewing" Bayesian updating and Bayesian consistency, both with and without the assumption of an absolute conviction that the distribution of what one sees over time is iid.

6.2.1. *The finite case.* Let S be a finite set, $S^\infty = \times_{t=1}^\infty$ the countable product of S. For any $t \geq 1$, let $h_t = (x_1, \ldots, x_t)$ be a point in S^t, and $A(h_t)$ the cylinder set determined by h_t,

$$A(h_t) = \{s : (z_1(s), \ldots, z_t(s) = (x_1, \ldots, x_t)\}.$$

For any $m \in \Delta(S)$, let m^∞ denote the distribution on S^∞ defined by

$$m^\infty(A(h_t) = \Pi_{n=1}^t m(x_n),$$

that is, m^∞ is the distribution of an infinite sequence of iid draws distributed according to m. Let $\lambda \in \Delta(S)$ denote the true distribution governing an iid set of draws, and let $\mu \in \Delta(\Delta(S))$ denote a prior distribution over the possible λ's. With beliefs μ, the prior probability that h_t happens is

$$Pr_\mu(h_t) := \int_{\Delta(S)} m^\infty(A(h_t))\, d\mu(m).$$

Definition 6.3. *For any Borel P on the csm (X, d), the* **support of** *P is* $\text{supp}(P) = \cap\{F : F$ *is closed and* $P(F) = 1\}$*, the smallest closed set having probability* 1.

Having a large support set means that a probability is "everywhere." The following, the proof of which uses only additivity and the fact that a set is closed iff its complement is open, is meant to indicate why this is a sensible interpretation.

Lemma 6.4. $\text{supp}(\mu) = X$ *iff for all non-empty, open G, $\mu(G) > 0$.*

Homework 6.6. *If* $\text{supp}(\mu) = \Delta(S)$*, then for all t and all h_t, $Pr_\mu(h_t) > 0$.*

Don't get too excited by the previous result, if $\mu = \delta_G$ and $G(s) > 0$ for all s, then for all t and all h_t, $Pr_\mu(h_t) > 0$. We need $Pr_\mu(h_t) > 0$ in order to use Bayes law to update beliefs after every possible partial history h_t.

After seeing h_t, the prior beliefs μ_t are updated to $\mu_t(\cdot|h_t)$, defined by

$$\mu_t(E|h_t) = \frac{\int_E m^\infty(A(h_t))\,d\mu(m)}{\int_{\Delta(S)} m^\infty(A(h_t))\,d\mu(m)} = \frac{\int_E m^\infty(A(h_t))\,d\mu(m)}{Pr_\mu(h_t)}.$$

Definition 6.5. *The beliefs-truth pair* (μ, λ) *is* **consistent** *if*

$$\lambda^\infty\{s : \lim_t \rho_w(\mu_t(\cdot|A(z_1(s), \dots, z_t(s))), \lambda) = 0\} = 1,$$

that is, almost always, Bayesian updating leads to the truth.

It is true (but not as easy to prove as it should be) that if μ is full support, then for all λ, (μ, λ) is consistent. When we look in $\Delta(S)$, the set of full support distributions is "most" of the set of distributions. In this sense, consistency is generic. However, even for consistent beliefs-truth pairs, the convergence can be awfully slow.

Homework 6.7. *Suppose that* $S = \{H, T\}$ *so that* $\Delta(S) = [0, 1]$*, with* $x \in [0, 1]$ *giving the probability of* H*. Suppose that* $\mu \in \Delta([0, 1])$ *has the cdf* $F_\mu(x) = x^r$*. Suppose that* λ *corresponds to* $x = 0$*, that is, to* T *with probability* 1*.*

1. *Find, as a function of* r*, the rate at which* $\rho_w(\mu_t, \lambda) \to 0$*. [Intuitively, for* r *large, the convergence should be very slow.]*
2. *Suppose that* μ *is replaced by a probability* ν *having the properties that* $\nu(\mathbb{Q}) = 1$*, for all* $q \in \mathbb{Q} \cap [0, 1]$*,* $\nu(q) > 0$*, and for all* $x \in [0, 1]$*,* $F_\nu(x) \leq F_\mu(x)$*. Show that* (ν, λ) *is consistent.*

If beliefs are not full support, consistency may fail.

Homework 6.8. *Suppose that* $S = \{H, T\}$ *so that* $\Delta(S) = [0, 1]$*, with* $x \in [0, 1]$ *giving the probability of* H*. Suppose that for some* $0 < s < 1$*,* $\mu \in \Delta([0, 1])$ *has the cdf*

$$F_\mu(x) = \begin{cases} 0 & \text{if } x \leq s \\ (x - s)^r/(1 - s)^r & \text{if } s < x \leq 1 \end{cases}$$

Show that for all t *and all* h_t*,* $Pr_\mu(h_t) > 0$*. Nevertheless, if* λ *is given by any* $x \in [0, s)$*, then the pair* (μ, λ) *is not consistent.*

6.2.2. *The infinite case.* The calculations we've done so far leaned pretty heavily on the iid assumption. This can be reformulated as the assumption that we are interested in updating to distributions over S^∞ that are in a very small subset of $\Delta(S^\infty, \mathcal{C}^\circ)$. The general question of what distributions, λ, in $\Delta(S^\infty, \mathcal{C}^\circ)$ are learnable is the topic of [13], which produces, whenever possible, an asymptotic Bayesian representation of λ by setting $\mu(\cdot) = \lambda(\cdot|\mathcal{F}_\infty)$. It seems pretty clear that this induces a pretty special, non-generic, relation between beliefs, μ, and the truth, λ, in order to get at learnability, which is something like consistency. In fact, we saw if λ picks one of a set of iid probabilities, then $\mu(\cdot) = \lambda(\cdot|\mathcal{F}_\infty)$ gives exactly that representation, and learnability and consistency are identical.

One can still ask about consistency in the context of infinite metric spaces. For the simplest starting point, one would like to know how widespread consistency is when $S = \mathbb{N}$ and the iid assumption is in place. It turns out that the full support assumption is no longer sufficient. Intuitively, this is plausible because we could get arbitrarily slow convergence in the finite case (Homework 6.7), and getting the slowest of an infinite sequence of slower and slower convergences might get us no convergence at all.

Borel probabilities μ on a metric space (X, d) are said to have **full support** if $\text{supp}(\mu) = X$. We're about to use the following, fairly immediate consequence of Lemma 6.4.

Lemma 6.6. *If X' is a countable dense subset of X and $\mu(x') > 0$ for all $x' \in X'$, then* $\text{supp}(\mu) = X$.

Another useful fact is that for the metric space $(\mathbb{N}, d)$ and G_n, G Borel probabilities on $\mathbb{N}$, $\rho_w(G_n, G) \to 0$ iff for all finite $E \subset \mathbb{N}$, $G_n(E) \to G(E)$.

Homework 6.9. *This problem consists of some preliminaries and then a proof that there is a dense set of full support beliefs, denoted here by μ_ϵ, with the property that for every λ in a dense subset of $\Delta(\mathbb{N})$, the pair (μ_ϵ, λ) is* **not** *consistent.*

1. *Let* $M_n \subset \Delta(\mathbb{N})$ *be the set of probability distributions,* P*, with* $\#\text{supp}(P) = n$ *and* $P(m) \in \mathbb{Q}$ *for all* $m \in \mathbb{N}$*.* $M' = \cup_n M_n$ *is a countable dense subset of* $\Delta(\mathbb{N})$*.*
2. *The set* Δ^{fs} *of full support probabilities is dense in* $\Delta(\mathbb{N})$*.*
3. *The set* Δ^{fs} *is dense in itself, that is, for any* $G \in \Delta^{fs}$*, the set* $\Delta^{fs} \setminus \{G\}$ *is dense in* Δ^{fs}*, hence dense in* $\Delta(\mathbb{N})$*.*
4. *Let* $\nu \in \Delta(\Delta(\mathbb{N}))$ *satisfy* $\nu(M') = 1$ *and for all* $P \in M'$*,* $\nu(P) > 0$*. Let* $G \in \Delta^{fs}$*. For any* $\epsilon \in (0, 1)$*, define* $\mu_\epsilon \in \Delta(\Delta(\mathbb{N}))$ *by* $\mu_\epsilon = (1 - \epsilon)\nu + \epsilon\delta_G$*. For all* $\epsilon \in (0, 1)$*,* $\text{supp}(\mu_\epsilon) = \Delta(\Delta(\mathbb{N}))$*.*
5. *For any* λ *in the dense set* $\Delta^{fs} \setminus \{G\}$*, every pair* (μ_ϵ, λ) *fails consistency.*
6. *The set of* μ_ϵ *constructed as above is dense in* $\Delta(\Delta(\mathbb{N}))$*.*

6.3. **Conjugate families and fictitious play.** In the case that observations are iid $\lambda \in \Delta(X)$, beliefs, μ, are points in $\Delta(\Delta(X))$. A class of priors, $M_\Theta = \{\mu_\theta : \theta \in \Theta\}$, is a **conjugate family** if each $\mu_t(\cdot|h_t) \in M_\Theta$. For general csm X, setting $M_\Theta = \Delta(\Delta(X))$ gives a conjugate family, one that is generally too big to be useful. For finite X, taking $M_\Theta = \{\delta_\lambda\}$ when $\text{supp}(\lambda) = X$ gives another conjugate family, one too small to be useful unless the truth is actually λ.

The typical conjugate families have $\Theta \subset \mathbb{R}^\ell$ for some $\ell \in \mathbb{N}$. For example, for each $r \in \mathbb{R}$, let $\lambda_r = N(r, \sigma^2)$ for some fixed $\sigma^2 > 0$. Let $\Theta = \mathbb{R}^1$, and for each $\theta \in \Theta$, have $\mu_\theta \in \Delta(\Delta(\mathbb{R})$ be described by picking a λ_r where $r \sim N(\theta, \psi^2)$ for some fixed $\psi^2 > 0$. In words, ones beliefs about λ is that they are normal with variance σ^2 and unknown mean, and that one's prior about the mean is that it is distributed $N(\theta, \psi^2)$. Having beliefs like that and updating according to Bayes rule leads to well-known statistical procedures.

Mis-specification of the model/beliefs is a severe problem with classes of distributions that are parametrized by finite dimensional vectors. Slightly more formally, when S is infinite, $\Delta(S)$ is a convex subset of an infinite dimensional vector space. This means that $\Delta(\Delta(S))$ is "even more" infinite dimensional. If $\theta \mapsto \mu_\theta$ is a smooth mapping from a finite dimensional Θ to $\Delta(\Delta(S))$, one cannot expect M_Θ to be a large or representative subset. One can prove that M_Θ is what is called a "shy" subset of $\Delta(\Delta(S))$, and that there is a shy subset E of $\Delta(S)$ with the property that $\mu_\theta(E) = 1$ for all $\theta \in \Theta$. Being a "shy" subset is the infinite dimensional analogue of a "Lebesgue null" set. This means that typical conjugate families do not cover anything but a very small subset of $\Delta(S)$.

Anyhow, all the generalities aside, the class of Dirichlet distributions form a conjugate family for multinomial sampling, and Bayesian updating looks just like the fictitious play updating of the γ_t^i. Therefore, if we believe that all the players' beliefs about others' behavior is that they are iid according to some distribution, $p \in \Delta(S_{-i})$, and our beliefs about p are Dirichlet, then Bayesian updating is exactly the same as forming the γ_t as the convex combination of the empirical D_t and using those beliefs as the new parameters of the Dirichlet. While this is nice, it may well have nothing to do with how the people are actually behaving, and, since the people never abandon their priors (even after several thousand cycles), it's not a generally attractive model of behavior.

7. Some "Evolutionary" Dynamics and "Evolutionarily" Stable Strategies

In this section, we're going to look at dynamics in which strategies that do better are played a higher proportion of the time. This can be a story about one person's likelihood of playing a given strategy, as in Hart and Mas-Colell's [10] work on the convergence to correlated equilibria. Usually, however, it is a story with evolutionary overtones to it, that is, a story about a large population of people/creatures where the population average number of times a strategy is played increases with the payoff to the strategy. This is what gives the work an "evolutionary" flavor.

This could be done in discrete time, and sometimes is, but we'll follow the tradition and use continuous time and differential equations to specify the dynamic systems. Closely related to the dynamics is the idea of an Evolutionarily Stable Strategy (ESS), which gives a (sometimes empty) subset of the Nash equilibria.

An essential difference between the types of dynamic stories, and an essential limitation on most of the work that's been done in this part of the field is the assumption that there is only one population of creatures interacting with other members of the same population. This means that the theory only addresses symmetric games, a very small subset of the games we might care about. We'll start with these one population dynamics, then go to a famous predator prey two population example, then look at a variety of other examples and applications.

7.1. **ESS and the one population model.** Here's the class of games to which these solution concepts apply.

Definition 7.1. *A two person game* $\Gamma = (S_i, u_i)_{i=1,2}$ *is* **symmetric** *if*

1. $S_1 = S_2 = S = \{1, 2, \dots, n, \dots, N\}$,
2. *for all* $n, m \in S$, $u_1(n, m) = u_2(m, n)$

We have a big population of players, typically $\Omega = [0, 1]$, we pick 2 of them independently and at random, label them 1 and 2 **but do not tell them the labels**, and they pick $s_1, s_2 \in S$, then they receive the vector of utilities $(u_1(s_1, s_2), u_2(s_1, s_2))$. It is very important, and we will come back to this, that the players do not have any say in who they will be matched to.

Let p_n be the proportion of the population picking strategy $n \in S$, and let $\sigma = (\sigma_1, \dots, \sigma_N) \in \Delta(S)$ be the summary statistic for the population propensities to play different strategies. This summary statistic can arise in two ways: **monomorphically**, i.e. each player ω plays the same σ; or **polymorphically**, i.e. a fraction σ_n of the population plays pure strategy n. (There is some technical mumbo jumbo to go through at this point about having uncountably many independent choices of strategy in the monomorphic case, but I know both nonstandard analysis and some other ways around this problem.)

In either the monomorphic or the polymorphic case, a player's expected payoff to playing m when the summary statistic is σ is

$$u(m,\sigma) = \sum_{n \in S} u(m,n)\sigma_n,$$

and their payoff to playing $\tau \in \Delta(S)$ is

$$u(\tau,\sigma) = \sum_{m \in S} \tau_m u(m,\sigma) = \sum_{m,n \in S} \tau_m u(m,n)\sigma_n.$$

From this pair of equations, if we pick a player at random when the population summary statistic is σ, the expected payoff that they will receive is $u(\sigma,\sigma)$.

Now suppose that we replace a fraction ϵ of the population with a "mutant" who plays m, assuming that $\sigma \neq \delta_m$. The new summary statistic for the population is $\tau = (1-\epsilon)\sigma + \epsilon\delta_m$. Picking a non-mutant at random, their expected payoff is

$$v^{\epsilon}_{n-m} = u(\sigma,\tau) = (1-\epsilon)u(\sigma,\sigma) + \epsilon u(\sigma,\delta_m).$$

Picking a mutant at random, their expected payoff is

$$v^{\epsilon}_{m} = u(m,\tau) = (1-\epsilon)u(m,\sigma) + \epsilon u(m,m).$$

Definition 7.2. *A strategy σ is an* **evolutionarily stable strategy (ESS)** *if there exists an $\underline{\epsilon} > 0$ such that for all $\epsilon \in (0,\underline{\epsilon})$, $v^{\epsilon}_{n-m} > v^{\epsilon}_{m}$.*

An interpretation: a strategy is an ESS so long as scarce mutants cannot succesfully invade. This interpretation identifies success with high payoffs, behind this is the idea that successful strategies replicate themselves. In principle this could happen through inheritance governed by genes or through imitation by organisms markedly more clever than (say) amœbæ.

Homework 7.1. *The following three conditions are equivalent:*

1. *σ is an ESS.*
2. *For all $\tau \neq \sigma$, $u(\sigma,\sigma) > u(\tau,\sigma)$ or $u(\sigma,\sigma) = u(\tau,\sigma)$ and $u(\sigma,m) > u(m,m)$.*
3. *$(\exists \underline{\epsilon} > 0)(\forall \tau \in B(\sigma,\epsilon)\ \tau \neq \sigma)[u(\sigma,\tau) > u(\tau,\tau)]$.*

The last condition and the compactness of $\Delta(S)$ imply that there is at most a finite number of ESS's. It also, more seriously, implies that in extensive form games, where there are often connected sets of equilibria, none of the connected sets can contain an ESS. This means that applying this kind of evolutionary argument to extensive form games is going to require some additional work. We're probably not going to have the time to do it though.

Since mutants are supposed to be scarce, we might expect them to play pure strategies. In the polymorphic interpretation of play, this is all that they could do. One might believe that the geometry of the simplex and convex combinations imply that we can replace mutants playing pure strategies δ_m by mutants playing any mixed strategy $\tau \neq \sigma$. This is not true. This means that, in some contexts, there may be a serious evolutionary advantage to being able to randomize. However, since the example is non-generic, the succeeding problem means you should take this conclusion with a grain of salt.

Homework 7.2. *The first strategy in the following game is an ESS if only pure strategy mutants are allowed, but a mixed strategy mutant playing* $(0, \frac{1}{2}, \frac{1}{2})$ *can succesfully invade.*

		Player 2		
		1	2	3
	1	$(1,1)$	$(1,1)$	$(1,1)$
Player 1	2	$(1,1)$	$(0,0)$	$(3,3)$
	3	$(1,1)$	$(3,3)$	$(0,0)$

Homework 7.3. *If σ is an ESS, then σ is a Nash equilibrium, if σ is a strict Nash equilibrium, then σ is an ESS.*

The following game may be familiar to you, if not, it should be, it's about an important set of ideas and it shows that ESS's need not exist: The E-Bay auction for a Doggie-shaped vase of a particularly vile shade of green has just ended. Now the winner should send the seller the money and the seller should send the winner the vile vase. If both act honorably, the utilities are $(u_b, u_s) = (1, 1)$, if the buyer acts honorably and the seller dishonorably, the utilities are $(u_b, u_s) = (-2, 2)$, if the reverse, the utilities are $(u_b, u_s) = (2, -2)$, and if both act dishonorably, the utilities are $(u_b, u_s) = (-1, -1)$.

For a (utility) cost s, $0 < s < 1$, the buyer and the seller can mail their obligations to a third party intermediary that will hold the payment until the vase arrives or hold the vase until the payment arrives, mail them on to the correct parties if both arrive, and return the vase or the money to the correct party if one side acts dishonorably. Thus, each person has three choices, send to the intermediary, honorable, dishonorable. The payoff matrix for the symmetric, 3×3 game just described is

		Seller		
		Intermed.	Honorable	Dishonorable
	Intermed.	1-s , 1-s	1-s , 1	-s , 0
Buyer	Honorable	1 , 1-s	1 , 1	-2 , 2
	Dishonorable	0 , -s	2 , -2	-1 , -1

Homework 7.4. *Verify the following: despite the labelling of the players by distinct economic roles, the game is symmetric; the game has a unique, full support mixed strategy equilibrium; the unique mixed strategy equilibrium is invadable by Honorable mutants* [*use the second equivalent formulation of ESS's*]*; therefore the game has no ESS* [*since every ESS is Nash*].

7.2. **ESS without blind matching.** ESS's do not tell players who they are matched against. It is blind. We're going to spend a little bit of time looking at what happens if we remove the blindness aspect a little bit (I learned to think about these issues from reading [24]).

7.2.1. *Breaking symmetry.* The starting point is the following game shows that the symmetry assumption has some real bite.

	1	2
1	$(0,0)$	$(2,2)$
2	$(2,2)$	$(0,0)$

Homework 7.5. *Show that* $(\frac{1}{2}, \frac{1}{2})$ *is the unique ESS for the game just given.*

We're now going to look at what happens if matching is no longer blind, but is subject to evolutionary pressures.

Let's suppose that mutants arise who can mess with the rules of the game. Specifically, suppose that there are mutants who can condition on some aspect of the meeting, in effect, allowing them to condition on whether they are player 1 or player 2. Suppose these mutants played the strategy "have s match which player I am." When a non-mutant meets either a non-mutant or a mutant, they receive expected utility of 1. When a mutant meets a non-mutant, they get an expected utility of 1, when they meet another mutant, they get the higher expected utility of 2. This invasion works. This strongly suggests that evolutionary pressures will push toward assortative matching, at least, in this game.

7.2.2. *Cycles of invasion and processing capacity.* Here's another example that pushes our thinking in another direction.

	1	2
1	$(3,3)$	$(7,1)$
2	$(1,7)$	$(5,5)$

You should recognize this as a version of the Prisoners' Dilemma. It has a unique strict equilibrium, hence a unique ESS. Continuing in the "messing with the rules of the game" vein, let us suppose that mutants arise who can recognize each other, and they play the strategy 1 if playing a non-mutant, 2 if playing a mutant. When a non-mutant meets a mutant or a non-mutant, they will get utility of 3, when a mutant meets a non-mutant, they will receive a utility of 3, when they meet another mutant, they will receive a utility of 5. Again, an invasion that works.

Let us now suppose that the mutants of the previous paragraph have taken over. Remember, they have this vestigial capacity to recognize the previous population of amœbæ that were playing 1. Now suppose a new strain of mutant arises, mutant′, one that cannot be distinguished from the present population by the present population, but that plays 1 unless they meet another mutant′, in which case they play 2. Again, this invasion is successful. One can imagine such cycles continuing indefinitely. There are other variants. Suppose mutant″ arise that cannot be distinguished from mutant′, but which plays the strategy 1 all the time. They can succesfully invade up to some proportion of the population, at which point they and the population of mutant′ are doing equally well. That population is inavadable by mutant‴ who recognizes both of the previous types, plays 1 against all others who play 1, plays 2 against itself and against all others who play 2.

What I like about this arms race is that it shows how there may be reproductive advantages to having more processing capacity, and that we expect there to be cycles of behavior.

7.2.3. *Cheap talk as a substitute for non-blind matching.* Consider the coordination game

	a	b
a	$(2,2)$	$(-100,0)$
b	$(0,-100)$	$(1,1)$

Homework 7.6. *The two strict equilibria of this game are ESS's, but the mixed Nash equilibrium is not an ESS.*

The (b,b) ESS risk-dominates the (a,a) ESS, even though (a,a) Pareto dominates (b,b). Suppose that we add a first, communicative stage to this game, a stage in which the two players simultaneously announce a message $m \in \{\alpha,\beta\}$ and can condition play in the second period on the results of the first stage. We assume that the talk stage is cheap, that is,

1. any conditioning strategy for second period play is allowed, and

2. utility is unaffected by messages.

Communication does not improve things using regular old equilibrium analysis.

Homework 7.7. *In the extensive form game just described, the set of proper equilibria contains all the Nash plays of the second stage game. (The same is true for stable sets of equilibria, but that's a bit harder).*

Homework 7.8. *Consider the (proper) equilibrium in which the players say "α" and play "b" no matter what is said in the first period. That is, the equilibrium is a set of "liars" who ignore communication. This is not an ESS, it is invadable by mutants who say "β," play a if there are two β's, and otherwise play b, that is, by mutants who "lie," but pay attention to communication.*

This seems to suggest that evolutionary pressures could hitch a ride on the possible efficiency gains of communication. It's not quite true, the ESS for the first stage of this game is unique: it involves each message being sent with equal probability. In the second stage, the messages are ignored and then either efficient or inefficient pure strategies are played. These are called "babbling" equilibria, in these equilibria, what people say is "full of sound and fury, signifying nothing."

Now suppose that the inefficient communication ESS was being played, and musically talented mutants come along who pitch their voices in a subtle fashion not recognized by the existing population, and, if they run into each other, play the efficient equilibrium. Essentially, the present population has tuned out the messages, the mutants invent, and use, a new message. (Sometimes, this new message is called a "secret handshake.") Again, we can see cycles coming into being, but can conclude that inefficient play will be invaded by talkative, i.e. communicative, mutants.

7.2.4. *Morals from ESSs without blind matching.* We could reformulate any I-person game into a symmetric game played by one population simply by picking I individuals at a time, telling them their role, and then giving them the payoffs $u_i(s)$ when they are in role i and $s \in S$ is picked. This would mean that each organism (or whatever) would need to have (coded in their genes) instructions on what to play in every role they might come into. This seems a bit of a stretch, and we'll not go down that road.

The various examples above showed that there can be advantages to being able to tell what kind of person you're matched with, that blindness may not be adaptive. Information flows are crucial, and we must think carefully about the informational flow assumptions that we make. This may take us to cycles or arms races. If we had a dynamic, or class of dynamics, that we trusted, this would not be a major intellectual concern, we'd simply follow the dynamics. This would involve analyzing comparative dynamics rather than comparative statics, and this is harder, but not fundamentally horrible. Still, before going to the evolutionary dynamics, let's look at multiple population versions of ESSs.

7.3. **ESS and the multiple population model.** Let $\Gamma = (S_i, u_i)_{i \in I}$ be an i person game, $\sigma = (\sigma_i)_{i \in I}$ a strategy for Γ. The idea now is that each $i \in I$ is drawn from a population Ω_i and matched against an independ set of draws form the populations Ω_j, $j \neq i$. Mutants are supposed to be rare, so let us imagine that ϵ of them happen to one of the populations, the idea being that the probability of mutants happening in two of the population pools would be on the order of ϵ^2 and we're dealing with small ϵ's. Above, σ was an ESS if no small enough proportion of mutants can invade and change the payoffs so that the extant population is doing less well. Now that we have many populations, we want no mutant invasion of i to upset either the optimality of population i's distribution or the optimality of population j's distribution, $j \neq i$.

Suppose the population summary statistic is $\sigma^* = (\sigma_i^*, \sigma_{-i}^*)$. After population i is invaded by mutants playing $m \in S_i$, the population summary statistic is (τ_i, σ_{-i}^*) where $\tau_i(\epsilon) = (1-\epsilon)\sigma_i^* + \epsilon\delta_m$.

Definition 7.3. σ^* *is a* **multi-population ESS** *if for all i and all $m \in S_i$, there exists an $\underline{\epsilon} > 0$ such that for all $\epsilon \in (0, \underline{\epsilon})$,*

1. $u_i(\sigma^*) > u_i(\tau_i(\epsilon), \sigma_{-i}^*)$, *and*
2. *for all $j \in I$, $u_j(\sigma^*) \geq u_j(\tau_i(\epsilon), \sigma_i^*)$.*

Note that every multipopulation ESS is a Nash equilibrium, by the second line.

An interpretation: a strategy is an ESS so long as scarce mutants in population i cannot succesfully invade population i, and the presence of mutants in population i does not affect the optimality of the population(s) $j \neq i$. Again, this interpretation identifies success with high payoffs, behind this is the idea that successful strategies replicate themselves.

Let us suppose that we are dealing with a generic game in the sense that there are finitely many equilibria, and at each of the finitely many equilibria, σ^*, i's choice matters: for all $\sigma^* \in Eq(\Gamma)$, there exists an $m \in S_i$ such that for all $j \in I$, the vector $(\partial u_j(\tau_i(\epsilon), s_{-i})/\partial \epsilon)_{|\epsilon=0} \in \mathbb{R}^{S_{-i}}$ has no zero components and no equal components.

Lemma 7.4. *In a game where i's choice matters, no strategy involving mixing is an ESS.*

Proof: Suppose that σ^* is an equilibrium in such a game and suppose that some $j \in I$ is playing strategy not at a vertex. In this case, when population i is invaded by mutants playing m, the j's utility to playing the actions in the support of σ_j^* move at different rates over any interval $(0, \underline{\epsilon})$. Therefore, the mixed strategy is no longer optimal and mutants will invade population j. ∎

7.4. **"Evolutionary" Differential Equations.** We'll start with some of the simplest differential equations, hopefully this will be a reminder, but if not, it's supposed to be your introduction. After this, we'll do a famous two population model, the Lotka-Volterra predator/prey model. Then we'll go back to the symmetric games in which we discussed ESS's and look at what are called "monotone" dynamics, the famous "replicator" dynamics are a special case.

7.4.1. *The simplest two cases.* We imagine that a "state" variable, $x \in \mathbb{R}^n$, moves (?evolves?) over time in a smooth way. This means that $t \mapsto x(t)$ is differentiable. We use $\dot{x}$ and dx/dt and $D_t x$ for the derivative of the time path $t \mapsto x(t)$. What is sneaky about differential equations and related models is the (brilliant) simplifying assumption that $\dot{x}$ is a function of the state, and sometimes of the point in time too,

$$\dot{x} = f(x), \text{ or } \dot{x} = f(x, t).$$

By way of parallel, the first, $\dot{x} = f(x)$, is like a stationary Markov chain, while the second, $\dot{x} = f(x,t)$, is like a Markov chain with transition probabilities that vary over time.

The second simplest differential equation ever invented is

$$\dot{x} = rx \in \mathbb{R}^1.$$

The class of solutions to it is $x(t) = be^{rt}$ for some constant b. If we specify the value of x at some point in time we will have nailed down the behavior, $x(0) = x_0$ is the usual convention for naming the time and place. This is exponential growth or exponential decay.

The next step, if we're thinking about populations is to introduce carrying capacities. For example, suppose that the "carrying capacity" of an environmental niche is γ, the equation might well be something like

$$\dot{x} = r\left(1 - \frac{x}{\gamma}\right)x.$$

Notice that solving for $\dot{x} = 0$ gives either $x = 0$ or $x = \gamma$, extinction, or right at carrying capacity. Before solving this, note that

$$sgn\,(\dot{x}) = sgn\,(\gamma - x) = sgn\,\left(1 - \frac{x}{\gamma}\right).$$

So, when $x > \gamma$, i.e. the population is above the carrying capacity, the population declines, and when below, it increases. By doing some algebra,

$$x(t) = \frac{\gamma}{Be^{-rt} + 1}$$

for some constant B determined by $x(0) = x_0$. Specifically, $B = (\gamma - x_0)/x_0$.

7.4.2. *Lotka-Volterra.* Remember the famous movie line, "I have always trusted the kindness of strangers"?

There are two types of prey, those who trust in the kindness of strangers, and those who carry deadly force. There are two kinds of strangers, the kind that are trustworthy and the preying kind. Let x be the fraction of trusting prey, and y the fraction of preying strangers. The preying strangers grow at a rate $\delta_1 x$ and are sent to meet their maker at a rate $\gamma_1(1-x)$. From this

$$\frac{\dot{y}}{y} = \delta_1 x - \gamma_1(1-x),$$

equivalently,

$$\dot{y} = \delta y(x-\gamma),$$

where $\delta = \delta_1 + \gamma_1 > 0$ and $\gamma = \gamma_1/(\delta_1+\gamma_1) \in (0,1)$.

Suppose that x following the differential equation

$$\dot{x} = x(g-\mu y),$$

where g is the growth rate of prey, $\mu > 0$, and μy is the rate at which the prey is removed by the predators.

Solve for $\dot{x} = \dot{y} = 0$, draw a phase diagram. An explicit solution to the system of equations is not known. We could simulate it and watch the trajectories. This is tempting in the age of the computer. However, a trajectory is a set

$$T(x_0,y_0) = \{(x(t),y(t)) : \ t \geq 0, \ \dot{x} = x(g-\mu y), \ \ \dot{y} = \delta y(x-\gamma), \ (x(0),y(0)) = (x_0,y_0)\}.$$

It would be nice to say something about the shape of the sets $T(x_0,y_0)$. Let's look for

$$S = \{(x,y) : \frac{dy}{dx} = \frac{\delta y(x-\gamma)}{x(g-\mu y)}\}.$$

If we can get an expression for the corresponding set of x and y, up to some constant say, then we're pretty sure we've got a function which is constant over the sets $T(x_0,y_0)$. Rearrange so the y's and x's are on separate sides, integrate both, and we get the expression $y^g x^{\delta\gamma} e^{-(\mu y+\delta x)} = e^C$. That is, we expect that

$$S = \{(x,y) : y^g x^{\delta\gamma} e^{-(\mu y+\delta x)} = e^C\}$$

for some constant C. It's now merely a tedious check that along any trajectory of the system of differential equations, the expression holds as an equality. Now we do something really tricky: along the ray from the origin

$$x = \gamma s, \ y = (g/\mu)s,$$

we have an expression of the form

$$se^{-s} = D$$

for some constant D. When $s = 1$, we're at the stationary point of the system. This is strictly decreasing in s for $s > 1$ and increasing for $s < 1$, hence the orbits of the sytem are closed.

Whew!

7.4.3. *Monotone dynamics.* We just saw that multiple populations can be analyzed, but that it's complicated. Compare the result about "ESS" for multiple populations, only the strict equilibria were possible. However, perhaps we end up with sensible looking dynamics. Let $\sigma(t)$ be the population summary statistic at time t.

Monotone dynamics come in many flavors.

First, for all $\sigma_i \gg 0$, if $u_i(\sigma(t), s_i) > (=) u_i(\sigma(t), t_i)$, then

$$\frac{\dot{\sigma}_i(s_i)}{\sigma_i(s_i)} > (=) \frac{\dot{\sigma}_i(t_i)}{\sigma_i(t_i)}.$$

Second, we could apply the previous to mixed strategies: for all $\sigma_i \gg 0$, if $u_i(\sigma(t), \sigma_i) > (=) u_i(\sigma(t), \tau_i)$, then

$$\sum_{s_i \in S_i} (\sigma_i(s_i) - \tau_i(s_i)) \frac{\dot{\sigma}_i(s_i)}{\sigma_i(s_i)} > (=) 0.$$

Third, for all $\sigma_i \gg 0$, sign agreement.

Fourth, inner product > 0.

References

[1] Abreu, Dilip (1988): "OSPC," *Econometrica*,

[2] Bergin, James. E'trica on the frailty of convergence results.

[3] Billingsley, Patrick. *Probability and Measure.*

[4] Blackwell, David (19??): "Approachability,"

[5] Blackwell, David and Lester Dubins (1962): "Merging of Opinions with Increasing Information," *Annals of Mathematical Statistics* **38**, 882-886.

[6] Chu, James Chia-Shang, Maxwell Stinchcombe, and Halbert White (1996): "Monitoring Structural Change," *Econometrica* **64**(5), 1045-1065.

[7] Fudenberg, Drew and David Kreps (1988): "Learning, experimentation and equilibrium in games," photocopy, Department of Economics, Stanford University.

[8] Fudenberg, Drew and David Levine (19??): "Limit games and limit equilibria," *Journal of Economic Theory*

[9] Fudenberg, Drew and David Levine (1998): *The Theory of Learning in Games*. Cambridge: MIT Press.

[10] Hart, Sergiu and Andreu Mas-Colell. Convergence to correlated eq'a and the Blackwell approachability article they're based on.

[11] Hillas, John (1990): "On the Definition of the Strategic Stability of Equilibria," *Econometrica* **58**, 1365-1390.

[12] Ichiishi, Tatsuro (1983): *Game theory for economic analysis*. New York : Academic Press.

[13] Jackson, Matthew, Ehud Kalai, and Rann Smorodinsky (1999): "Bayesian Representation of Stochastic Processes Under Learning: De Finetti Revisited,“ *Econometrica* **67**(4), 875-893.

[14] Kalai, Ehud and Ehud Lehrer (1993): "Rational Learning Leads to Nash Equilibrium," *Econometrica* **61**(5), 1019-1046.

[15] Kalai, Ehud and Ehud Lehrer (1993): "Subjective Equilibrium in Repeated Games," *Econometrica* **61**(5), 1231-1240.

[16] Kandori, M., George Mailath, and Rafael Rob (1993): "Learning, Mutation, and Long Run Equilibrium," *Econometrica* **61**, 27-56.

[17] Myerson, R. (1978): "Refinement of the Nash Equilibrium Concept," *International Journal of Game Theory* **7**, 73-80.

[18] Nachbar, John (1997): "Prediction, Optimization, and Learning in Repeated Games," *Econometrica*, **65**(2), 275-309.

[19] Nelson, Edward (1987): *Radically Elementary Probability Theory*, Annals of mathematics studies no. 117. Princeton, N.J. : Princeton University Press.

[20] Samuelson, Larry (19??): Either his book or some article(s).

[21] Pollard, David (1984): *Convergence of Stochastic Processes*. New York: Springer-Verlag.

[22] Selten, R. (1975): Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games, *International Journal of Game Theory* **4**, 25-55.

[23] Simon, Leo and Maxwell Stinchcombe (1995): "Equilibrium Refinement for Infinite Normal Form Games," *Econometrica* **63**(6), 1421-1444.

[24] Skyrms, Brian (199?): *Evolution of the Social Contract*.

[25] Stinchcombe, Maxwell (1997): "Countably Additive Subjective Probabilities," *Review of Economic Studies* **64**, 125-146.

[26] Stinchcombe, Maxwell (1990): "Bayesian Information Topologies," *Journal of Mathematical Economics* **19**, 3, 233-254.

[27] Stinchcombe, Maxwell (1993): "A Further Note on Bayesian Information Topologies," *Journal of Mathematical Economics* **22**, 189-193.

[28] Young, H. Peyton (1998): *Individual strategy and social structure: an evolutionary theory of institutions*. Princeton: Princeton University Press.