ADVANCED MICRO: ECONOMICS 387L.26, FALL 2013

MAXWELL B. STINCHCOMBE

1. Putting *'s on Everything

The readings for the first two sections are Ch. 11.1 - 11.2 in Corbae, D., Stinchcombe, M. B., and Zeman, J. (2009). *An introduction to mathematical analysis for economic theory and econometrics*. Princeton University Press, Princeton, NJ, Ch. 8.3 - 8.4 in Fudenberg, D. and Tirole, J. (1991). *Game theory*. MIT Press, Cambridge, MA, and Lindstrøm, T. (1988). An invitation to nonstandard analysis. In *Nonstandard analysis and its applications (Hull, 1986)*, volume 10 of *London Math. Soc. Stud. Texts*, pages 1–105. Cambridge Univ. Press, Cambridge.

1.1. Properties of Purely Finitely Additive Point Masses.

The basic device for us is the set of μ equivalence classes of sequences where μ is a purely finitely additive "point mass." We will later show that there exists a "probability" on the integers, μ, with the following properties:

1. for all $A \subset \mathbb{N}$, $\mu(A) = 0$ or $\mu(A) = 1$;
2. $\mu(A \cup B) = \mu(A) + \mu(B)$ for all disjoint $A, B \subset \mathbb{N}$;
3. $\mu(\mathbb{N}) = 1$; and
4. $\mu(A) = 0$ if $A \subset \mathbb{N}$ is finite.

Some useful and pretty obvious consequence of these properties:

1. If $E_1, \ldots, E_K$ is a partition of $\mathbb{N}$, then $\mu(E_k) = 1$ for exactly 1 of the partition elements. To give a formal argument, start from the observation that this is true if $K = 2$ (from the second property above), and if true for K, then it is true for K, then it is true for $K + 1$.
2. If $\mu(A) = \mu(B) = 1$, then $\mu(A \cap B) = 1$. Since $\mu(A \cup B) = 1$ because $A \subset (A \cup B)$, this consequence follows from the observation that $\mu(A^c) = \mu(B^c) = 0$ so that $A \setminus B = A \cap B^c$ is a subset of a 0 set, hence has mass 0, and, by the same reasoning, $B \setminus A$ is a null set. Finally, $A \cup B$ is the disjoint union of the sets $(A \setminus B)$, $(B \setminus A)$, and $(A \cap B)$.

1.2. The Equivalence Classes.

For any set X, $X^{\mathbb{N}}$ denotes the set of sequences in X. We define two sequences $x = (x_1, x_2, \ldots)$ and $y = (y_1, y_2, \ldots)$ to be equivalent, $x \sim_\mu y$, if $\mu(\{n \in \mathbb{N} : x_n = y_n\}) = 1$. By the second of the consequences just given, this is an equivalence relation. For any $x \in X^{\mathbb{N}}$, $\langle x_1, x_2, x_3, \ldots \rangle$ denotes the equivalence class of x.

We define "star X," written ${}^*X$ to be the set of all equivalence classes, ${}^*X = (X^{\mathbb{N}})/\sim_\mu$. This gives us new objects to use. The pattern is to "put *'s on everything," where by 'everything' we mean relations, functions, sets, classes of sets, correspondences, etc.

Example 1.2.1. *$^*[0,1]$ contains the equivalence class $dt := \langle 1, \frac{1}{2}, \frac{1}{3}, \ldots\rangle$, called a **non-standard** number as well as all of the equivalence classes $r := \langle r, r, r, \ldots\rangle$, and these are called the standard numbers. Since $\mu(\{n \in \mathbb{N} : \frac{1}{n} < r\}) = 1$ if $r \in (0,1]$ and $\mu(\{n \in \mathbb{N} : 0 < \frac{1}{n}\}) = 1$, we write $0^* <dt^* < r$. This means that our new number, dt, is strictly greater than 0 and strictly less than all of the usual, standard strictly positive numbers. We write this as $dt \simeq 0$ and say that dt **is infinitesimal**. The only standard number in $^*[0,1]$ that is infinitesimal is 0.*

Example 1.2.2. *Another infinitesimal is $dx = \langle 1, \frac{1}{4}, \frac{1}{9}, \ldots\langle$, indeed, $dx = (dt)^2$ and $0 < dx < dt < r$ (where we have not put *'s on the less than signs). Yet another infinitesimal is $dy = \langle \frac{1}{10}, \frac{1}{10^2}, \frac{1}{10^3}, \ldots\rangle$. Now $0 < dy < dx < dt < r$, and $s := \frac{dx}{dy} = \langle \frac{1/n}{1/10^n}\rangle = \langle 10, \frac{10^2}{2}, \frac{10^3}{3}, \ldots\rangle$ has the property that for any $R \in \mathbb{R}$, $R < s$. We either say that s is an **unlimited number** or we say that it is an **infinite number**.*

Example 1.2.3. *For $x, y \in {}^*\mathbb{R}$, we define $x-y = \langle x_1-y_1, x_2-y_2, x_3-y_3, \ldots\rangle$, $x+y = \langle x_1+y_1, x_2+y_2, x_3+y_3, \ldots\rangle$, $x \cdot y = \langle x_1 \cdot y_1, x_2 \cdot y_2, x_3 \cdot y_3, \ldots\rangle$, $|x| = \langle |x_1|, |x_2|, |x_3|, \ldots\rangle$, and so on. We write that $x \simeq y$ if $|x - y| \simeq 0$, and say that x and y are infinitely close to each other, or we say that they are at an infinitesimal distance from each other.*

Example 1.2.4. *A function $f : [0,1] \to \mathbb{R}$ is continuous iff for all $a \in [0,1]$, $[x_n \to a] \Rightarrow [f(x_n) \to f(a)]$. For any $x \in {}^*[0,1]$, we define $^*f(x) = \langle f(x_1), f(x_2), f(x_3), \ldots\rangle$. From this, you can see that the function f is continuous at a iff $[x \simeq a] \Rightarrow [^*f(x) \simeq f(a)]$. An infinitesimal move in the domain of the function leads to an infinitesimal move in the range.*

1.3. **Some Real Analysis Problems.** The due date for these problems is Mon. Sept. 23. Hints come directly after.

The first set of problems are closely related to the previous examples.

A. CSZ 11.1.2.
B. CSZ 11.1.5.
C. CSZ 11.1.7.
D. CSZ 11.1.9.
E. CSZ 11.1.11.

The next set of problems ask you to push yourself further through the patterns of "putting *'s on everything."

F. $n \mapsto s_n$ is a Cauchy sequence in $\mathbb{R}$ iff $^*s_n \simeq {}^*s_m$ for all $n, m \in {}^*\mathbb{N} \setminus \mathbb{N}$.
G. The continuous functions on $[0,1]$ are denoted $C([0,1])$, the metric we use on them is $d_\infty(f,g) = \max_{t\in[0,1]} |f(t) - g(t)|$.
   1. A function $f : [0,1] \to \mathbb{R}$ belongs to $C([0,1])$ iff for all $t_1 \simeq t_2 \in {}^*[0,1]$, $^*f(t_1) \simeq {}^*f(t_2)$.
   2. If $T \in {}^*\mathcal{P}_F([0,1])$, $^*d_H(T, {}^*[0,1]) \simeq 0$, and $t \in T$ solves $^*\max_{t\in T} {}^*f(t)$ for $f \in C([0,1])$, then $^\circ t$ solves $\max_{t\in[0,1]} f(t)$.
   3. Suppose that $f \in C([0,1])$ and that $f(0) > 0 > f(1)$. Using a set T as in the previous problem, show that $f(c) = 0$ for some $c \in (0,1)$.
H. Some exercises with derivatives and related. Throughout, $dx \neq 0$.

1. For $r \in \mathbb{R}$, we define $e^r = \sum_{n=0}^{\infty} \frac{r^n}{n!}$. Show that if $m, m' \in {}^*\mathbb{N} \setminus \mathbb{N}$ and $r \in {}^*\mathbb{R}$ is finite, then $\sum_{n=0}^{m} \frac{r^n}{n!} \simeq \sum_{n=0}^{m'} \frac{r^n}{n!}$.
2. Show that if $dx \simeq 0$, then $e^{dx} \simeq 1$ and $(e^{dx} - 1)/dx \simeq 1$. From this show that for any $x \in \mathbb{R}$, $\frac{e^{x+dx} - e^x}{dx} \simeq e^x$.
3. Show that for $x \in \mathbb{R}$ and $dx \simeq 0$, $\frac{(x+dx)^n - x^n}{dx} \simeq nx^{n-1}$.
4. If f and g are continuously differentiable at 0, $g'(0) \neq 0$, and $f(0) = g(0) = 0$, then $\lim_{x \to 0} \frac{f(x)}{g(x)} \simeq \frac{f(dx)}{g(dx)} \simeq \frac{f'(0)}{g'(0)}$.

I. Show that if $h \in {}^*\mathbb{R}_+ \setminus \mathbb{R}_+$, then $(\sqrt{h+1} - \sqrt{h}) \simeq 0$. From this conclude that $\lim_{x \to \infty}(\sqrt{x+1} - \sqrt{x}) = 0$.

J. For every $r \in \mathbb{R}$, there exists $q \in {}^*\mathbb{Q}$ such that ${}^\circ q = r$. In particular, $\{{}^\circ q : q \in {}^*\mathbb{Q}, q \text{ finite}\}$, is much larger than $\mathbb{Q}$, while $\{{}^\circ r : r \in {}^*\mathbb{R}, r \text{ finite}\} = \mathbb{R}$.

1.4. **Some Hints. Cauchy sequences**: A sequence in X is a function $s : \mathbb{N} \to X$, denoted above as $n \mapsto s_n$. The *'d version of a function is what one uses to think about the values of s_n, s_m for infinite m, m', that is $m, m' \in {}^*\mathbb{N} \setminus \mathbb{N}$. For a given sequence $n \mapsto s_n$, and $M \in \mathbb{N}$, define $\delta_M = \sup\{d(x_m, x_{m'}) : m, m' \geq M\}$. Note that $\delta_{M+1} \leq \delta_M$, and that the sequence is Cauchy iff $\delta_M \downarrow 0$.

- Therefore, for arbitrary $\epsilon \in \mathbb{R}_{++}$, ${}^*\{M \in \mathbb{N}\delta_M < \epsilon\}$ contains all infinite m, m' when $n \mapsto s_n$ is Cauchy. This means that for any $\epsilon > 0$ and any pair of infinite integers m, m', $d(s_m, s_{m'}) < \epsilon$, i.e. $d(s_m, s_{m'}) \simeq 0$.
- Now suppose that ${}^*s_n \simeq {}^*s_m$ for all $n, m \in {}^*\mathbb{N} \setminus \mathbb{N}$. For arbitrary $\epsilon \in \mathbb{R}_{++}$, the internal set $\{M \in {}^*\mathbb{N} : (\forall m, m' \geq M)[d(s_m, s_{m'}) < \epsilon]\}$ contains arbitrary small infinite elements, hence contains finite elements.

Continuous functions: By definition, a function $f : [0,1] \to \mathbb{R}$ is continuous at $a \in [0,1]$ if for all sequences $x_n \to a$, $f(x_n) \to f(a)$.

- Let $x = \langle x_1, x_2, \ldots \rangle$ be the equivalence class of any sequence converging to a, for any $\epsilon \in \mathbb{R}_{++}$, $\{n \in \mathbb{N} : d(f(x_n), f(a)) < \epsilon\}$ has only a finite complement, hence $d({}^*f(x), f(a)) \simeq 0$.
- Since $[0,1]$ is compact, for any $t_1 \simeq t_2 \in {}^*[0,1]$, there is a unique $a \in [0,1]$ such that $a = {}^\circ t_1 = {}^\circ t_2$, and $d({}^*f(t_1), f(a)) \simeq 0$ and $d({}^*f(t_2), f(a)) \simeq 0$.
- If $T \in {}^*\mathcal{P}_F([0,1])$ with $d_H(T, {}^*[0,1]) \simeq 0$, and $f : [0,1] \to \mathbb{R}$ is continuous and $f(0) > 0 > f(1)$, consider the internal set $T_{++} = \{t \in T : {}^*f(t) > 0\}$, set $t' = \max T_{++}$. Let $a = {}^\circ t'$ and note that $f(a) = 0$.

Theorem 1.1 (Robinson). *A metric space (X, d) is compact iff for every $x \in {}^*X$, there exists an $a \in X$ such that $d(a, x) \simeq 0$.*

Proof. Recall that a metric space (X, d) is compact iff it is both totally bounded and complete.
$\Rightarrow$: Let $x = \langle x_1, x_2, \ldots \rangle \in {}^*X$ with X complete and totally bounded. By total boundedness, there exists a finite subset $F_1 = \{a_{1,1}, \ldots, a_{1,M_1}\}$ such that for all $a \in X$, $d(a, F_1) < \frac{1}{2^1}$. Disjointify the finite open cover of X given by $\{B_{1/2^1}(a_{1,m} : m \leq M_1\}$ into the sets $A_{1,m}$. The sets $E_{1,m} := \{n \in \mathbb{N} : x_n \in A_{1,m}\}$ partition $\mathbb{N}$, hence exactly one, say E_{1,m_1}, of them has μ-mass 1. Let n_1 be the first element in E_{1,m_1}.

Disjointify an open cover of A_{1,m_1} by $\frac{1}{2^2}$ balls and repeat, letting n_2 be the first element of the set of integers E_{2,m_2} that has μ-mass 1.

Continuing gives a Cauchy subsequence x_{n_k}, which, by completeness has a limit in X, call it a. For every $k \in \mathbb{N}$, $d(a,x) < \frac{1}{2^k}$, hence $d(a,x) \simeq 0$.

$\Leftarrow$: If X is not totally bounded, then there exists $\epsilon \in \mathbb{R}_{++}$ such that for every finite set F, there exists an $x \in X$ such that $d(x,F) \geq \epsilon$. Let $F_1 = \{x_1\}$, pick x_{n+1} such that $d(x_{n+1}, F_n) \geq \epsilon$, and set $F_{n+1} = F_n \cup \{x_{n+1}\}$. Let $x \in {}^*X = \langle x_1, x_2, \ldots \rangle$. There can be no $a \in X$ that is the standard part of x because this would mean that $d(a,x) < \epsilon/3$, which would imply that along a subsequence x_{n_k}, $d(x_{n_k}, x_{n_{k'}}) < d(x_{n_k}, a) + d(a, x_{n_{k'}}) < \epsilon/3 + \epsilon/3 < \epsilon$.

If X is not complete, then there exists a Cauchy sequence x_n that is not converging to any $a \in X$. Let $x \in {}^*X$ be the equivalence class $\langle x_1, x_2, \ldots \rangle$. If $a \in X$ is the standard part of x, then $d(x_{n_k}, a) \to 0$ for some subsequence, but if any subsequence of a Cauchy sequence converges, the whole sequence converges. □

2. Some Equilibrium Refinement

2.1. **Notation.** A finite game is $\Gamma = (A_i, u_i)_{i \in I}$ where $A := \times_{i \in I} A_i$ is finite and $u_i \in \mathbb{R}^A$. Mixed strategies for $i \in I$ and $\Delta(A_i) := \{\mu_i \in \mathbb{R}_+^{A_i} : \sum_{a_i \in A_i} \mu_i(a_i) = 1\}$. Utilities are extended to $\times_{i \in I} \Delta(A_i)$ by $u_i(\mu) = \sum_{a \in A} u_i(a) \Pi_{i \in I} \mu_i(a_i)$. The (relative) interior of $\Delta(A_i)$ is denoted Δ_i° and defined by $\Delta_i^\circ = \{\mu_i \in \Delta(A_i) : \mu_i \gg 0\}$. We will use the notation $\mu \setminus \nu_i$ for the vector $(\mu_1, \ldots, \mu_{i-1}, \nu_i, \mu_{i+1}, \ldots, \mu_I)$ and we will pass back and forth between point mass on a_i, i.e. δ_{a_i}, and a_i as convenient.

For $\mu \in \times_{i \in I} \Delta(A_i)$ and $j \in I$, $Br_j(\mu) := \text{argmax}_{a_i \in A_i} u_i(\mu \setminus a_i)$. With this notation we have the starting point for non-cooperative game theory.

Definition 2.1. *μ^* is a* ***Nash equilibrium*** *if $(\forall i \in I)[\mu_i(Br_i(\mu)) = 1]$. The set of Nash equilibria for a game is denoted $Eq(\Gamma)$.*

2.2. **Refinement.** Especially when the strategies $\Delta(A_i)$ are the agent normal form strategies for an extensive form game, there are many Nash equilibria. One way to get rid of them is to ask that they be robust to infinitesimal perturbations in the games.

Here are three perturbation based equilibrium refinement concepts, in increasingly order of strength. After these three we have a version of a set-valued solution concept.

Definition 2.2. *For $\epsilon \in {}^*\mathbb{R}_{++}$, $\mu \in \times_{i \in I} {}^*\Delta_i^\circ$ is* ***ϵ-perfect*** *if*

$$(\forall i \in I)(\forall b_i \in A_i)[\,[\max_{a_i \in A_i} u_i(\mu \setminus a_i) > u_i(\mu \setminus b_i)] \Rightarrow [\mu_i(b_i) < \epsilon]\,]. \tag{1}$$

$\mu^* \in \times_{i \in I} \Delta(A_i)$ is a ***perfect equilibrium*** *if $\mu^* = {}^\circ\mu$ for some ϵ-perfect μ with $\epsilon \simeq 0$. The set of perfect equilibria for a game is denoted $Per(\Gamma)$.*

Definition 2.3. *For $\epsilon \in {}^*\mathbb{R}_{++}$, $\mu \in \times_{i \in I} {}^*\Delta_i^\circ$ is* ***ϵ-proper*** *if*

$$(\forall i \in I)(\forall a_i, b_i \in A_i)[\,[u_i(\mu \setminus a_i) > u_i(\mu \setminus b_i)] \Rightarrow [\mu_i(b_i) < \epsilon \cdot \mu_i(a_i)]\,]. \tag{2}$$

$\mu^* \in \times_{i \in I} \Delta(A_i)$ is a ***proper equilibrium*** *if $\mu^* = {}^\circ\mu$ for an ϵ-proper μ with $\epsilon \simeq 0$. The set of proper equilibria for a game is denoted $Pro(\Gamma)$.*

Definition 2.4. $\mu^* \in \times_{i\in I}\Delta(A_i)$ *is* ***strictly perfect*** *if for all* $\mu \in \times_{i\in I}{}^*\Delta_i^\circ$,

$$[\,[\mu \simeq \mu^*] \Rightarrow (\forall i \in I)[\mu_i(Br_i(\mu) \simeq 1]\,]. \tag{3}$$

The set of proper equilibria for a game is denoted $Str(\Gamma)$.

Definition 2.5. *A closed connected set* $S \subset Eq(\Gamma)$ *is* ***robust to perturbations*** *if*

$$(\forall \mu \in \times_{i\in I}{}^*\Delta_i^\circ)[\,[d_H(\mu, {}^*S) \simeq 0] \Rightarrow (\forall i \in I)[\mu_i(Br_i(\mu)) \simeq 1]\,]. \tag{4}$$

A closed and connected $S \subset Eq(\Gamma)$ *is* ***p-stable*** *if it is robust to perturbations and no closed, connected strict subset of* S *is robust to perturbations.*

2.3. **Some Game Theory Problems.**

J. Prove the following inclusion results.
   1. Every perfect equilibrium is a Nash equilibrium, $Per(\Gamma) \subset Eq(\Gamma)$.
   2. Every proper equilibrium is a perfect equilibrium, $Pro(\Gamma) \subset Per(\Gamma)$.
   3. $Pro(\Gamma) \neq \emptyset$.
   4. Every strictly perfect equilibrium is a perfect equilibrium, $Str(\Gamma) \subset Per(\Gamma)$.
   5. Every strictly perfect equilibrium is a proper equilibrium, $Str(\Gamma) \subset Pro(\Gamma)$.
   6. If S is a p-stable set, then $S \subset Per(\Gamma)$.
   7. If S is a p-stable set, then $S \cap Pro(\Gamma) \neq \emptyset$.

K. If $\mu^* \in Per(\Gamma)$, then there exists $\mu \in {}^* \times_{i\in I} \Delta_i^\circ$ such that $(\forall i \in I)[\mu_i(Br_i(\mu^*) \simeq 1]$, but the reverse is not true. [This captures the difference between sequential and trembling hand perfect equilibria.]

L. Give the set of perfect equilibria for the following game and show that it strictly contains the set of proper equilibria.

	L	R	A_2
T	$(1,1)$	$(0,0)$	$(-1,-2)$
B	$(0,0)$	$(0,0)$	$(0,-2)$
A_2	$(-2,-1)$	$(-2,0)$	$(-2,-2)$

M. The following game has no strictly perfect equilibrium. Find its p-stable set of equilibria.

	L	M	R
T	$(1,2)$	$(1,0)$	$(0,0)$
B	$(1,2)$	$(0,0)$	$(1,0)$

3. Decision Theory with Full Support Probabilities

Readings for this section are Blume, L., Brandenburger, A., and Dekel, E. (1991a). Lexicographic probabilities and choice under uncertainty. *Econometrica*, 59(1):61–79 and Blume, L., Brandenburger, A., and Dekel, E. (1991b). Lexicographic probabilities and equilibrium refinements. *Econometrica*, 59(1):81–98.

Looking at strategies in ${}^*\Delta_i^\circ$ made equilibrium refinement work pretty well, essentially because at all points in a game tree, the players had to pay attention to all possibilities, but could assign relatively small probability to non-best responses. Another aspect of strictly positive probabilities is that one never has to condition on a null set, all of the conditional probabilities are well-defined. This will save us a great deal of hassle once we get to stochastic process theory.

3.1. **The Basic Model.** We assume that we have a probability space $(\Omega, \mathcal{F}, P)$ where Ω is a *-finite set, $\mathcal{F} = \mathcal{P}(\Omega)$, and P, the **prior distribution**, is strictly positive. Utility depends on a random state, $\omega \in \Omega$, and the choice of action, $a \in A$, $u(a, \omega)$. When we work with games, ω will be the choices of other players.

Let $E_1, \dots, E_K$ be *-finite partition of Ω, representing what the decision maker will know before their decision. That is, before making a decision, one learns which E_k, $k \in \{1, \dots, K\}$ contains ω. In extensive form games, this corresponds to learning what information set we are at. Because P is strictly positive, we never divide by 0 in the following observation,

$$(\forall A \in \mathcal{F}) \left[P(A) = \sum_k P(E_k) \frac{P(A \cap E_k)}{P(E_k)} = \sum_k P(E_k) P(A|E_k) \right]. \tag{5}$$

Another way to put this is that one's **posterior beliefs**, that is, beliefs after having observed your information, about an event A are $P(A|E_k)$. This equation tells us that your average belief is your prior belief. As it holds for all A, we could write it $P(\cdot) = \sum_k P(\cdot|E_k) P_k$ in $\Delta(\Omega)$ where $P_k = P(E_k)$.

3.2. **Bridge Crossing.** The decision problem is

$$P : \quad \max_{a_1, \dots, a_K \in A} \int 1_{E_k} u(a_k, \omega)\, dP(\omega). \tag{6}$$

This kind of decision theory leads us to Bayes' law updating, and the K problems

$$P_k : \quad \max_{a \in A} \int u(a, \omega)\, dP(\omega|E_k). \tag{7}$$

Recall the saying, "I'll cross that bridge when I get to it." It is usually understood to mean that I'll figure out what I need to do once I know more about the decision problem. Here, what you will know is some one of the E_k.

Lemma 3.1 (Bridge-Crossing). *$(a_1^*, \dots, a_K^*)$ solves the decision problem P if and only if each a_k^* solves problem P_k.*

The Bridge-Crossing Lemma tells us that solving each P_k and putting it back together is the same as solving P, and vice versa. Defining $P(\cdot|E_k)$ when $P(E_k) = 0$ is not a straightforward business.

3.3. **Heirarchies of Beliefs.** For the rest of the semester, we will almost exclusively be looking at the case when $P \in {}^*\Delta^\circ(\Omega)$. In this case, $P(E_k) > 0$ for all k, which is nice. The difference between the infinitesimal and non-infinitesimal $P(\omega|E_k)$ gives rise to **heirachies** of beliefs as follows:

1. For $P \in {}^*\Delta^\circ(\Omega)$, let $Q_1 = {}^\circ P$, and let $E_1 = \{\omega : Q_1(\omega) > 0\}$.
2. If $E_1^c \neq \emptyset$, define $Q_2 = {}^\circ P(\cdot|E_1^c)$, and let $E_2 = \{\omega : Q_2(\omega) > 0\}$.
3. If $(E_1 \cup E_2)^c \neq \emptyset$, define $Q_3 = {}^\circ P(\cdot|(E_1 \cup E_2)^c)$, and let $E_3 = \{\omega : Q_3(\omega) > 0\}$.

4. And so on and so forth until some Q_K is reached (the process must end because Ω is finite).
5. The heirarchy associated with P is $(Q_1, \ldots, Q_K)$.

What is at work is the "order" of the infinitesimals.

Example 3.3.1. *Let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_7\}$, for an infinitesimal non-zero ϵ, and let*

$$\begin{aligned} P &= (\tfrac{1}{2}, \tfrac{1}{2} - (\epsilon + \epsilon^2), \tfrac{1}{3}\epsilon, \tfrac{1}{2}\epsilon, \tfrac{1}{6}\epsilon, \tfrac{3}{4}\epsilon^2, \tfrac{1}{4}\epsilon^2) \text{ so that } K = 3 \text{ and} \\ Q_1 &= (\tfrac{1}{2}, \tfrac{1}{2}, 0, 0, 0, 0, 0) \\ Q_2 &= (0, 0, \tfrac{1}{3}, \tfrac{1}{2}, \tfrac{1}{6}, 0, 0) \\ Q_3 &= (0, 0, 0, 0, 0, \tfrac{3}{4}, \tfrac{1}{4}). \end{aligned}$$

The two papers for this section work out some of the implications and properties of a decision theory based on heirarchies like this. For game theory, what is at work is perturbations in beliefs of agents in an agent normal form, and perturbations must arise from other players playing strictly positive strategies.

4. Random Variables for Stochastic Process Theory

We'll begin with a useful dynamic optimization problem with no stochastics, then turn to the basics of random variables on *-finite probability spaces, then to stochastic processes, which are collections of random variables indexed by time, in our case, by a *-finite time set with infinitesimal increments. The readings for this part of the course are Chapters 1 - 8 of Nelson, E. (1987). *Radically elementary probability theory*, volume 117 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ.

4.1. A Deterministic, Dynamic Problem.

For the following, partition the time interval $[0, T]$ into N equal length parts,

$$[0, T] = [0, T \cdot \tfrac{1}{N}) \cup [T \cdot \tfrac{1}{N}, T \cdot \tfrac{2}{N}) \cup \cdots \cup [T \cdot \tfrac{N-2}{N}, T \cdot \tfrac{N-1}{N}) \cup [T \cdot \tfrac{N-1}{N}, T].$$

At the beginning of each sub-interval, $T_n := [T \cdot \frac{n-1}{N}, T \cdot \frac{n}{N})$, we will make a decision. That decision cannot be changed until the beginning of the next interval. We will be particularly interested in what happens when $N \in {}^*\mathbb{N} \setminus \mathbb{N}$ so that $\Delta = dt := T \cdot \frac{1}{N} \simeq 0$.

N. We need to produce a total amount B of a good by time T. In each sub-interval, if we produce at a rate r per unit of time, our cost per unit is linear in r, with slope c_1. The cost of storing an amount R is $c_2 \cdot R$ per unit of time we store it. Letting $\Delta = T \cdot \frac{1}{N}$ be the length of the subintervals, the problem is

$$\min_{r_1, \ldots, r_N} \sum_n \left[c_1 \cdot (r_n)^2 \cdot \Delta + c_2 \sum_{m<n} r_m \cdot \Delta \right] \text{ subject to } \sum_n r_n \cdot \Delta = B.$$

1. Characterize the solution to this minimiztion problem.
2. For infinite N, the function $f(t) = \sum_n r_n 1_{T_n}(t)$ is near-standard in $C([0, T])$ iff for all $t_1 \simeq t_2$, $f(t_1) \simeq f(t_2)$.
3. For infinite N, the solution $f^*(t) = \sum_n r_n^* 1_{T_n}(t)$ is near-standard in $C([0, 1])$.

The point of developing all of this is that the standard part of f^* solves the problem

$$\min \int_0^T (c_1[S'(t)]^2 + c_2 S(t))\, dt \text{ subject to } S(0) = 0,\, S(T) = B. \tag{8}$$

To show this, we are going to need the nonstandard theory of integration, which often turns out to be a theory of summation. Integration is a crucial thing that one does with random variables, so we leave dynamic optimization knowing, as Arnold says, "I'll be back."

4.2. **Random Variables.** Fix a *-finite probability space $(\Omega, \mathcal{F}, P)$ with $\mathcal{F} = \mathcal{P}(\Omega)$ and $P \in {}^*\Delta^\circ(\Omega)$.

Definition 4.1. *A random variable is a an element X of $\mathbb{R}^\Omega$.*

Example 4.2.1. $\Omega = \{1, \ldots, n\}$, $P(A) = \frac{\#A}{n}$. *It is often useful to take $n = m!$ for some infinite integer m. Define $X(k) \in {}^*[0,1]$ by $X(k) = \frac{k}{n}$. For any $0 \leq a < b \leq 1$, $P({}^\circ X \in (a, b]) \simeq (b - a)$, which looks like the uniform distribution. A useful fact: if (M, d) is any complete separable metric space and μ is any probability on M, then there exists a measurable $f : [0,1] \to M$ such that $\mu(E) = Unif(f^{-1}(E))$. We will see that this means that this probability space allows us to model (up to an infinitesimal), all probabilities on all complete separable metric spaces.*

Example 4.2.2. *Let $T = \{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\}$ be a *-finite set infinitely close to $^*[0,1]$, i.e. with n an infinite integer. Let $\Omega = \{-1, +1\}^T$ and define P so that the canonical projection mappings* $\text{proj}_t(\omega) := \omega_t$ *are an i.i.d. collection with $P(\omega_t = -1) = P(\omega_t = +1) = \frac{1}{2}$.*

From this, define $X(t, \omega)$ as follows: $X(0, \omega) \equiv 0$, $X(1, \omega) = \frac{1}{\sqrt{n}}\omega_1$, $X(2, \omega) = \frac{1}{\sqrt{n}}(\omega_1 + \omega_2)$, ..., $X(\frac{k}{n}, \omega) = \frac{1}{\sqrt{n}} \sum_{i=1}^{k} \omega_i$. This is a random walk model.

If $r \in (0,1]$ and $\frac{k}{n} \simeq r$, then $X(\frac{k}{n}, \cdot)$ is the sum of infinitely many i.i.d. random variables that have been scaled so that $\text{Var}(X(\frac{k}{n}, \cdot)) \simeq r$, *and the oldest (deMoivre) arguments for the central limit theorem should tell you that $X(\frac{k}{n}, \cdot)$ is infinitely close to being a Gaussian distribution. Further for $k < k' < k''$, the random increments, $(X(\frac{k'}{n}, \cdot) - X(\frac{k}{n}, \cdot))$ and $(X(\frac{k''}{n}, \cdot) - X(\frac{k'}{n}, \cdot))$ are independent. If you've seen a definition of a Brownian motion, this looks awfully close.*

Example 4.2.3. *Let $T = \{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\}$ be a *-finite set infinitely close to $^*[0,1]$ as before. Let $\Omega' = \{0,1\}^T$ and define Q so that thet canonical projection mappings* $\text{proj}_t(\omega') := \omega'_t$ *are an i.i.d. collection with $P(\omega'_t = 1) = \lambda dt$ where dt is the infinitesimal size of the incremental steps in the time set, and $\lambda \in \mathbb{R}_{++}$. Define $Y(0, \omega') \equiv 0$, $Y(\frac{1}{n}, \omega') = \omega'_1$, $Y(\frac{k}{n}, \omega') = \sum_{i \leq k} \omega'_i$.*

For $r \in (0,1]$ and $\frac{k}{n} \simeq 0$, $Y(\frac{k}{n}, \cdot)$ is infinitely close to having a Poisson(λr) distribution. Further, for $k < k' < k''$, the random increments, $(Y(\frac{k'}{n}, \cdot) - Y(\frac{k}{n}, \cdot))$ and $(Y(\frac{k''}{n}, \cdot) - Y(\frac{k'}{n}, \cdot))$ are independent. If you've seen a definition of a Poisson process, this looks awfully close.

Example 4.2.4. *We can glue the previous two examples as $\Omega \times \Omega'$ so that P and Q are independent. After doing that, we can define $Z(\frac{k}{n}, (\omega, \omega')) = X(\frac{k}{n}, \omega) + Y(\frac{k}{n}, \omega')$.*

The central limit theorem has two parts: the one you are most likely to be used to is like the X process, composed of infinitely many identical random pieces, all of them very small, indeed infinitesimal; the other one is like the Y process, it allows the largest of the infinitely many identical random pieces to have a non-infinitesimal probability of being far away from 0, that is, not infinitesimal. This is the beginnings of the study of infinitely divisible distributions and Levy processes.

We are going to be particularly interested in properties of the set of time paths that arise. In the last three examples, pick an ω, an ω', or a pair (ω, ω'). The time paths are the functions $t \mapsto X(t, \omega)$, $t \mapsto Y(t, \omega')$ and $t \mapsto Z(t, (\omega, \omega'))$. After taking care of the fact that T is a strict subset of ${}^*[0, 1]$, we will see that the X paths are nearstandard in $C([0, 1])$, and that the Y paths and the Z paths are nearstandard in $D([0, 1])$ (the cadlag paths with the Skorohod metric). In order to do this, we need to understand what continuity looks like when stretched onto a set such as T, what bounded fluctuations look like, and how infinite sums behave. This last will get us into the *-finite versions of integration theory.

A big part of the arguments behind these results is Ulam's theorem: every countably additive probability on a complete separable metric space is tight, i.e. for all $\epsilon > 0$, there is a compact K_ϵ carrying at least mass $1 - \epsilon$. This means that understanding probabilities on compact (subsets of) metric spaces is part of the background we need. Behind that is what is (mistakenly) called the Riesz representation theorem.

4.3. **Expectations, Norms, Inequalities.** Recall that we have in mind a *-finite probability space $(\Omega, \mathcal{F}, P)$ with P strictly positive.

For $A \in \mathcal{F}$, $P(A) = \sum_{\omega \in A} P(\omega)$.

For a random variable X, $E\,X := \sum_\omega X(\omega)P(\omega)$.

4.3.1. *Expectations of some classic functions.* Bilinear forms: for random variables X and Y, $E\,XY = E\,YX$, and $E\,XX > 0$ unless $X = 0$. The norm of X is defined as $\sqrt{E\,XX}$ and denote $\|X\|_2$.

The class of constant random variables is a linear subspace of $\mathbb{R}^\Omega$. The mapping $X \mapsto E\,X$ is orthogonal projection onto that subspace, and $\varepsilon_X := X - E\,X$ to the projection.

For random variables X and Y: $\text{Var}(X) := E\,(X - E\,X)^2 \;= E\,(\varepsilon_X)^2$ is the variance of X; $\sqrt{\text{Var}(X)}$ is the standard deviation of X; $\text{Cov}(X, Y) := E\,(X - E\,X)(Y - E\,Y)$ is the covariance of X and Y; and $\rho_{X,Y} := \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}$ is the correlation, also known as the cosine of the angle between the vectors X and Y.

4.3.2. *Some norms.* The L^p-norms, $p \in [1, \infty)$ are $\|X\|_p := (E\,|X|^p)^{1/p}$ and $\|X\|_\infty := \max_{\omega \in \Omega} |X(\omega)|$. Recall Jensen's inequality, for any convex $f : \mathbb{R} \to \mathbb{R}$,

$$f(\sum_\omega X(\omega)P(\omega)) \leq \sum_\omega f(X(\omega))P(\omega), \text{ equivalently} \tag{9}$$

$$f(E\,X) \leq E\,f(X), \tag{10}$$

provable from the definition of convexity and induction.

Lemma 4.1. *For all random variables X and $\infty \geq p > q \geq 1$, $\|X\|_p \geq \|X\|_q$.*

Proof. The case $p = \infty$ is immediate, so we suppose that $\infty > p$.

Suppose first that $p > q = 1$, from Jensen's inequality using the convex function $f(r) = r^p$ for $r \geq 0$ on the random variable $|X|$, we have $(E\,|X|)^p \leq E\,|X|^p$. Taking p'th roots on both sides, $E\,|X| \leq (E\,|X|^p)^{1/p} = \|X\|_p$.

For the last case, $p > q \geq 1$, using the convex function $f(r) = r^{p/q}$ on the random variable $|X|^q$, we have $(E\,|X|^q)^{p/q} \leq E(|X|^q)^{p/q} = E\,|X|^p$. Taking p'th roots on both sides, $(E\,|X|^q)^{1/q} \leq (E\,|X|^p)^{1/p}$, that is, $\|X\|_p \geq \|X\|_q$. □

This means that $p \mapsto \|X\|_p$ is an increasing function of $p \in [1, \infty)$, strictly increasing unless X is a constant random variable. We know that all bounded monotonic functions on subsets on $\mathbb{R}$ have a supremum. We now ask what that supremum is. Let ω_0 solve the problem $\max_\omega |X(\omega)|$. Because $\|X\|_p \geq (|X(\omega_0)|^p P(\omega_0))^{1/p} = \|X\|_\infty P(\omega_0)^{1/p}$ and $P(\omega_0)^{1/p} \uparrow 1$ as $p \uparrow \infty$, we have $\lim_{p\uparrow\infty} \|X\|_p = \|X\|_\infty$, hence $\lim_{p\uparrow\infty} \|X\|_p = \|X\|_\infty$.

4.3.3. *The triangle inequality for norms.* Recall that for vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^\ell$, $\boldsymbol{xy} = \cos(\theta)\sqrt{\boldsymbol{xx}}\sqrt{\boldsymbol{yy}}$. From this, one can conclude that $|\boldsymbol{xy}| \leq \sqrt{\boldsymbol{xx}}\sqrt{\boldsymbol{yy}}$, and that $\sqrt{\boldsymbol{xx}} = \max\{\boldsymbol{xy} : \sqrt{\boldsymbol{yy}} = 1\}$. From this, we find for any vectors $\boldsymbol{r}, \boldsymbol{s}$, $\sqrt{(\boldsymbol{r}+\boldsymbol{s})(\boldsymbol{r}+\boldsymbol{s})} \leq \sqrt{\boldsymbol{rr}} + \sqrt{\boldsymbol{ss}}$. This is the basis of the triangle inequality — take $\boldsymbol{r} = \boldsymbol{x} - \boldsymbol{y}$ and $\boldsymbol{s} = \boldsymbol{y} - \boldsymbol{z}$ and find that the distance between $\boldsymbol{x}$ and $\boldsymbol{z}$ is less than the sum of the distances between $\boldsymbol{x}$ and $\boldsymbol{y}$ and between $\boldsymbol{y}$ and $\boldsymbol{z}$, $\sqrt{(\boldsymbol{x}-\boldsymbol{z})(\boldsymbol{x}-\boldsymbol{z})} \leq \sqrt{(\boldsymbol{x}-\boldsymbol{y})(\boldsymbol{x}-\boldsymbol{y})} + \sqrt{(\boldsymbol{y}-\boldsymbol{z})(\boldsymbol{y}-\boldsymbol{z})}$. We are after the same triangle inequality result for the $\|\cdot\|_p$ norms. It is called Minkowski's inequality. The starting point is the following, notice the part where the conditions for equality versus strict inequality appear.

Lemma 4.2 (Hölder)**.** *For any random variables X, Y, and any $p \in (1, \infty)$, if $\frac{1}{p} + \frac{1}{q} = 1$, then $|E\,XY| \leq \|X\|_p \|Y\|_q$.*

Proof. If $X = 0$ or $Y = 0$, the inequality is satisfied. For the other cases, we can divide each $X(\omega)$ by $\kappa_x := \|X\|_p$ and each $Y(\omega)$ by $\kappa_y := \|Y\|_q$. After we have done that, the left-hand side is also divided by $\kappa_x\kappa_y$, and we have reduced the problem to showing that $\sum_\omega |X(\omega)Y(\omega)|P(\omega) \leq 1$ when we know that $\sum_\omega |X(\omega)|^p P(\omega) = \sum_\omega |Y(\omega)|^q P(\omega) = 1$. Here is an odd-looking observation that will make the argument go, $\frac{1}{p} + \frac{1}{q} = 1$ so that we need only show that

$$\sum_\omega |X(\omega)Y(\omega)|P(\omega) \leq \frac{1}{p}\sum_\omega |X(\omega)|^p P(\omega) + \frac{1}{q}\sum_\omega |Y(\omega)|^q P(\omega). \tag{11}$$

Since the logarithm strictly concave, for any non-zero pair $X(\omega), Y(\omega)$, we have

$$\log(\frac{1}{p}|X(\omega)|^p + \frac{1}{q}|Y(\omega)|^q) \geq \frac{1}{p}\log(|X(\omega)|^p) + \frac{1}{q}\log(|Y(\omega)|^q) \tag{12}$$

with equality iff $|X(\omega)|^p = |Y(\omega)|^q$. Now, $\frac{1}{p}\log(|X(\omega)|^p) + \frac{1}{q}\log(|Y(\omega)|^q) = \log(|X(\omega)Y(\omega)|)$, so we have

$$\log(\frac{1}{p}|X(\omega)|^p + \frac{1}{q}|Y(\omega)|^q) \geq \log(|X(\omega)Y(\omega)|). \tag{13}$$

Since the logarithm is strictly monotonic, this means that $\frac{1}{p}|X(\omega)|^p + \frac{1}{q}|Y(\omega)|^q \geq |X(\omega)Y(\omega)|$. Taking the probability weighted convex combination of these inequalities yields what we were after,

$$\frac{1}{p}\sum_{\omega} |X(\omega)|^p P(\omega) + \frac{1}{q}\sum_{\omega} |Y(\omega)|^q P(\omega) \geq \sum_{\omega} |X(\omega)Y(\omega)| P(\omega) \tag{14}$$

because now the inequality holds even if $X(\omega) = 0$ or $Y(\omega) = 0$. □

Let us return to the part of the proof where we said that we have "equality iff $|X(\omega)|^p = |Y(\omega)|^q$." In more detail, what we showed is that $\sum_i |X(\omega)Y(\omega)| \leq \|X\|_p \|Y\|_q$ with equality when, for each i, we have $X(\omega) = \mathsf{sgn}\,(Y(\omega))|Y(\omega)|^{q/p}$. Combining yields the following.

Lemma 4.3. *For each $X \in \mathbb{R}^\ell$, $\|X\|_p = \max_{\|Y\|_q = 1} E\, XY$.*

From which we have the triangle inequality for the $\|\cdot\|_p$-norms.

Lemma 4.4 (Minkowski)**.** *For any $R, S \in \mathbb{R}^\ell$ and any $p \in (1, \infty)$, $\|R + S\|_p \leq \|R\|_p + \|S\|_p$.*

Proof. Same logic as the vector case. □

4.3.4. *Chebyshev.* Or was that Tchebyshov? For $X \geq 0$ and $r > 0$, $X \geq r1_{X \geq r}$ for every ω, hence $E\, X \geq rE\, 1_{X \geq r}$, turning it around we have

$$P(X \geq r) \leq \frac{1}{r} E\, X. \tag{15}$$

For any random variable X, $r > 0$, and $p > 0$, $\{|X| > r\} = \{|X|^p > r^p\}$ so that

$$P(|X| \geq r) \leq \frac{1}{r^p} E\, |X|^p. \tag{16}$$

I've seen both of these (and some other forms) called **Chebyshev's inequality**.

4.4. **Vector Algebras of Random Variables.** Recall for a random variable X and non-empty A, $E\,(X|A) = \frac{1}{P(A)} \sum_{\omega \in A} X(\omega)P(\omega) = \frac{1}{P(A)} EX \cdot 1_A$. When $\mathcal{A} = \{A_1, \ldots, A_K\}$ is a partition of Ω, $E\,(X|\mathcal{A})$ is, *by definition*, the random variable $\sum_k E\,(X|A_k) \cdot 1_{A_k}(\omega)$. Letting $X = 1_B$, we have $E\,(X|A) = P(B|A)$ and $P(B|\mathcal{A}) = E\,(1_B|\mathcal{A})$. This is another random variable.

The trick with vector algebras of functions is that they always take the form $\text{span}\,\{1_{A_k} : k = 1, \ldots K\}$ where $\{A_1, \ldots, A_K\}$ is a partition of Ω. This will mean that conditional expectations are orthogonal projections.

We are going to abuse notation and also use $\mathcal{A} \subset \mathbb{R}^\Omega$ to be a vector algebra (of functions). Contrary to some usages, we will always assume that our vector algebras contain the constant functions.

Definition 4.2. *$\mathcal{A} \subset \mathbb{R}^\Omega$ is a* ***vector algebra*** *if for all $X, Y \in \mathcal{A}$ and all $\alpha, \beta \in \mathbb{R}$,*

a. $\alpha 1_\Omega \in \mathcal{A}$,
b. $\alpha X + \beta Y \in \mathcal{A}$, *and*
c. $XY \in \mathcal{A}$.

Atoms are, historically, the indisolubly small objects.

Definition 4.3. *An **atom** of a vector algebra $\mathcal{A}$ is a maximal event on which all elements of $\mathcal{A}$ are constant.*

For any atom A and $\omega \notin A$, let $Y \in \mathcal{A}$ have the property that $Y(\omega) \neq Y(A)$, define

$$X_\omega(\cdot) = \frac{Y(\cdot) - Y(\omega)}{Y(A) - Y(\omega)}. \tag{17}$$

Note that $X_\omega(\omega) = 0$, $X_\omega(A) = 1$, and $X_\omega \in \mathcal{A}$. Now consider the function

$$R(\cdot) = \Pi_{\omega \notin A} X_\omega(\cdot). \tag{18}$$

What we have is that $R = 1_A$. That is the hard part of the argument behind the following.

Lemma 4.5. *If $\mathcal{A}$ is a vector algebra and $\{A_1, \ldots, A_K\}$ is its collection of atoms, then $\mathcal{A} = \text{span}\left(\{1_{A_k} : k = 1, \ldots, K\}\right)$.*

The following is a nearly immediate corollary.

Lemma 4.6. *The mapping $X \mapsto E\left(X|\mathcal{A}\right)$ is orthogonal projection.*

4.5. **Stochastic Processes.** For this course, the time set, $T = \{t_0 < t_1 < \cdots < t_N\}$, will be a member of ${}^*\mathcal{P}_F(\mathbb{R})$ with the property that $(t_{n+1} - t_n) \simeq 0$ for $n = 1, \ldots, N$. We will often take $t_0 = 0$ and $t_N = 1$, another frequent option has t_N unlimited.

Definition 4.4. *A **stochastic process on** T is a function $\xi : T \to \mathbb{R}^\Omega$.*

It is often useful to think of this in the form $\xi : T \times \Omega \to \mathbb{R}$. Time paths are functions $t \mapsto \xi(t, \omega)$, $\omega \in \Omega$. Picking ω according to P gives us the random time path $\xi(\cdot, \omega)$.

We are going to need ways of talking about strong laws, central limit theorems, properties of time paths, and we are going to want these to be internal. At this point, it makes sense to go back and be a bit more clear about what we meant by "putting *'s on everything."

5. Putting *'s on Everything Redux

The basic device for us is the set of μ equivalence classes of sequences where μ is a purely finitely additive "point mass." This material is based on Ch. 11.5 in Corbae, D., Stinchcombe, M. B., and Zeman, J. (2009). *An introduction to mathematical analysis for economic theory and econometrics.* Princeton University Press, Princeton, NJ After this we turn to superstructures, then putting *'s on superstructures.

5.1. **Purely Finitely Additive Point Masses.** We are interested in a purely finitely additive probability $\mu : \mathcal{P}(\mathbb{N}) \to \{0, 1\}$. Probabilities taking on only the values 0 or 1 are best thought of as point masses, and we will return to the question "Point mass on what?" at some point later. These probabilities can also be understood as $\mu(A) = 1_\mathcal{F}(A)$ where $\mathcal{F} \subset \mathcal{P}(\mathbb{N})$ is a **free ultrafilter on the integers**, which contains a bunch of as-yet-undefined terms.

$\mathcal{F} \subset \mathcal{P}(\mathbb{N})$ is a **filter** if it is closed under finite intersections, $A, B \in \mathcal{F}$, and supersets, $A \subset B$ and $A \in \mathcal{F}$ imply $B \in \mathcal{F}$.

Examples: $\mathcal{F}(n) = \{A \in \mathcal{P}(\mathbb{N}) : n \in A\}$; the Frechet filter (aka the cofinite filter), $\mathcal{F}^{cof} = \{A \in \mathcal{P}(\mathbb{N}) : A^c \text{ is finite}\}$; the trivial filter, $\mathcal{F} = \{\mathbb{N}\}$; the largest filter, $\mathcal{F} = \mathcal{P}(\mathbb{N})$.

A filter is **proper** if it is a proper subset of $\mathcal{P}(\mathbb{N})$, so no proper filter can contain $\emptyset$. *We will only work with proper filters from here onward.*

Note that $\bigcap\{A : A \in \mathcal{F}(n)\} = \{n\} \neq \emptyset$ while $\bigcap\{A : A \in \mathcal{F}^{cof}\} = \emptyset$. A filter $\mathcal{F}$ is **free** if $\bigcap\{A : A \in \mathcal{F}\} = \emptyset$.

A (proper) filter is **maximal** if it is not contained in any other filter. A (proper) filter is an **ultrafilter** if for all $A \in \mathcal{P}(\mathbb{N})$, $A \in \mathcal{F}$ or $A^c \in \mathcal{F}$.

$\mathcal{F}(n)$ is an ultrafilter, and cannot be a strict subset of any other (proper) filter.

Lemma 5.1. *A (proper) filter is maximal iff it is an ultrafilter.*

Proof. A little bit of arguing. □

Since $\mathcal{F}^{cof}$ is a proper, free filter, the following implies that free ultrafilters exist, at least if you accept Zorn's Lemma, which is equivalent to the Axiom of Choice.

Theorem 5.1. *Every proper filter is contained in an ultrafilter.*

Proof. Zorn's lemma plus the previous result. □

Relevant properties of $\mu(A) := 1_{\mathcal{F}}(A)$ when $\mathcal{F}$ is a free ultrafilter: $\mu(A) = 0$ for all finite A; $\mu(A \cup B) = \mu(A) + \mu(B)$ if $A \cap B = \emptyset$; $\mu(\mathbb{N}) = 1$; $[\mu(A) = \mu(B) = 1] \Rightarrow [\mu(A \cap B) = 1]$; $\mu(A) = 1$ and $A \subset B$ imply $\mu(B) = 1$; if $A_1, \ldots, A_K$ is a partition of $\mathbb{N}$, then $\mu(A_k) = 1$ for exactly one $k \in \{1, \ldots, K\}$.

5.2. **The equivalence relation $\sim_\mu$ and $^*X$.** For any set X, $X^{\mathbb{N}}$ denotes the class of X-valued sequences. For $x, y \in X^{\mathbb{N}}$, $x \sim_\mu y$ if $\mu(\{n \in \mathbb{N} : x_n = y_n\}) = 1$. We define **star-**$X$ by $^*X := X^{\mathbb{N}}/\sim_\mu$.

We will spend the rest of the semester working out what we have defined, and what it is good for. Special cases of interest take $X = \mathbb{R}$, or $X = \mathcal{P}_F(A)$, the class of finite subset of a set A. To do all of this once in an consistent fashion, we work with superstructures.

5.3. **Superstructures.** Readings: Ch. 2.13 and 11.2 in Corbae, D., Stinchcombe, M. B., and Zeman, J. (2009). *An introduction to mathematical analysis for economic theory and econometrics.* Princeton University Press, Princeton, NJ

We start with a set S containing $\mathbb{R}$ and any other points we think we may need later (which will not be very much).

Definition 5.1. *Define $V_0(S) = S$ and $V_{n+1}(S) = V_n(S) \cup \mathcal{P}(S)$. The **superstructure over** S is $\cup_{n=0}^{\infty} V_n(S)$. For any $x \in V(S)$, the **rank** of x is the smallest n such that $x \in V_n(S)$. S is a **set**, and anything in $V(S)$ with rank 1 or higher is a **set**, nothing else is a set.*

In particular, every set has finite rank, which avoids Russell's paradox. A **statement** $\mathbb{A}(x)$ is the indicator function of set, where we interpret $\mathbb{A}(x) = 1$ as "the statement $\mathbb{A}$ is true for x."

Examples: ordered pairs; functions from $\mathbb{R}$ to $\mathbb{R}$; the set of sequences in $\mathbb{R}$; the set of Cauchy sequences in $\mathbb{R}$; $\mathbb{R}^\ell_+$; rational preference relations on $\mathbb{R}^\ell_+$; rational preferences on $\mathbb{R}^\ell_+$ that can be represented by C^∞ utility functions; the Hilbert cube $[0,1]^{\mathbb{N}}$ with the metric $d(x,y) = \sum \frac{|x_n - y_n|}{2^n}$; the collection of $\mathcal{G}_\delta$'s in the Hilbert cube; the collection of Polish spaces; the collection of compact metric space games with I players.

5.4. **Defining $V(^*S)$ inductively.** Now would be a good time to recall the properties of our $\{0,1\}$-valued, purely finitely additive μ.

1. Let G_n be a sequence in $V_0(S)$, define $(G_1, G_2, \ldots) \sim (H_1, H_2, \ldots)$ if $\mu(\{n \in \mathbb{N} : G_n = H_n\}) = 1$ and for any sequence, let $\langle G_1, G_2, \ldots \rangle$ denote its equivalence class. $V_0(^*(S)$ is defined as the set of these equivalence classes. If $G = \langle G, G, G, \ldots \rangle$, then G is a **standard point**, otherwise it is **nonstandard point**.
   a. $0 = \langle 0, 0, 0, \ldots \rangle$, more generally $r = \langle r, r, r, \ldots \rangle$, $r \in \mathbb{R}$, are typical standard points.
   b. $\langle 1, \frac{1}{2}, \frac{1}{3}, \ldots \rangle \simeq 0$, $\langle r+1, r+\frac{1}{2}, r+\frac{1}{3}, \ldots \rangle$, and $\langle 1, 4, 9, 16, 25, \ldots \rangle$ are nonstandard points, an infinitesimal, a near-standard (aka limited) point, and an infinite (aka unlimited) point.
2. Let G_n be a sequence in $V_1(S)$ that is *not* a sequence in $V_0(S)$. $V_1(^*S)$ is defined as the union of $V_0(^*S)$ and the set of μ-equivalence classes of such sequences. An element $x = \langle x_n \rangle$ of $V_0(^*S)$ belongs to $G = \langle G_n \rangle$ if $\mu\{n \in \mathbb{N} : x_n \in G_n\} = 1$, written $x^* \in G$ or $x \in G$. If $G = \langle G, G, G, \ldots \rangle$, then G is **standard**, otherwise it is **internal**.
   a. $\langle [0,1], [0,1], [0,1], \ldots \rangle$ is the standard set we denote $^*[0,1]$, $^*\mathbb{R}_+ = \langle \mathbb{R}_+, \mathbb{R}_+, \mathbb{R}_+, \ldots \rangle$. $^*[0,1]$ contains the standard point $\langle r, r, r, \ldots \rangle$ as long as $0 \leq r \leq 1$, $^*\mathbb{R}_+$ contain unlimited points such as the factorials $\langle n! \rangle$. $^*[0,1]$ also contains the infinitesimal $\langle 1, \frac{1}{2}, \frac{1}{3}, \ldots \rangle$, and the nearstandard point $\langle r+1, r+\frac{1}{2}, r+\frac{1}{3}, \ldots \rangle$ as long as $0 \leq r < 1$.
   b. $F = \langle \{0,1\}, \{0, \frac{1}{2}, 1\}, \{0, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, 1\}, \ldots$ is an internal set satisfying $d_H(F, {}^*[0,1]) = \langle \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots \rangle \simeq 0$. The function d_H does not belong to $V_1(^*S)$, and we should be able to figure out when it does appear.
3. Let G_n be a sequence in $V_{n+1}(S)$ that is **not** a sequence in $V_n(S)$. And so forth and so on . . . .
   a. The Hausdorff metric for $\mathbb{R}$ is a function from pairs of compact subsets of $\mathbb{R}$ to $\mathbb{R}_+$. Every compact subset of $\mathbb{R}$ belongs to $V_1(S)$. The class of compact sets belongs to $V_2(S)$. Every ordered triple of the form (K_1, K_2, r), $r \in \mathbb{R}$, belongs to $V_3(S)$. d_H is a particular subset of such triples, hence belongs to $V_4(S)$. Letting $\mathcal{K}_{\mathbb{R}}$ denote the compact subsets of $\mathbb{R}$, for every pair $K_a = \langle K_{a,1}, K_{a,2}, \ldots \rangle$ and $K_b = \langle K_{b,1}, K_{b,2}, \ldots \rangle$ in $^*\mathcal{K}_{\mathbb{R}}$, we have set things up so that $d_H(K_a, K_b) = \langle d_H(K_{a,1}, K_{b,1}), d_B(K_{a,2}, K_{b,2}), \ldots \rangle$.
   b. If $(\Omega, \mathcal{F}, P)$ is a finite probability space with $\mathcal{F} = \mathcal{P}(\Omega)$, then $\mathbb{R}^\Omega$ is the set of random variables on Ω. If $(\Omega, \mathcal{F}, P) = \langle (\Omega_1, \mathcal{F}_1, P_1), (\Omega_2, \mathcal{F}_2, P_2), (\Omega_3, \mathcal{F}_3, P_3) \ldots \rangle$, then $\mathbb{R}^\Omega = \langle \mathbb{R}^\Omega_1, \mathbb{R}^\Omega_2, \mathbb{R}^\Omega_3, \ldots \rangle$ is the set of *-random variables on the internal set $\Omega = \langle \Omega_1, \Omega_2, \Omega_3, \ldots \rangle$.

Some more examples.

Example 5.4.1. $^*\mathbb{N}$ *is standard while* $F = \langle\{k/2^n : k = 0, \ldots, n \cdot 2^n\}\rangle$ *is an internal subset of* $^*\mathbb{R}_+$ *with the property that for all limited* $r \in \mathbb{R}$, $d(r, F) \simeq 0$.

Example 5.4.2. $^*C([0,1])$ *is standard while* $Poly = \langle \text{span}(\{x^k : k = 0, \ldots n\})\rangle$ *is an internal subset of* $^*C([0,1])$*, and the Stone-Weierstrass theorem tells us that every standard* $f \in {}^*C([0,1])$, $d(f, Poly) \simeq 0$.

6. Internal Sets for Stochastic Processes

To recognize when we have an internal set, it is useful to know when we don't.

6.1. **Some External Sets.**

Theorem 6.1. *The following sets are external.*

a. $\{n \in {}^*\mathbb{N} : n \text{ is standard}\}$.
b. $\{n \in {}^*\mathbb{N} : n \text{ is nonstandard}\}$.
c. $\{r \in {}^*\mathbb{R} : r \text{ is limited}\}$.
d. $\{r \in {}^*\mathbb{R} : r \text{ is unlimited}\}$.
e. $\{r \in {}^*\mathbb{R} : r \text{ is infinitesimal}\}$.

Proof. □

Here is an implication that will be useful many times.

Lemma 6.1 (Robinson)**.** *If* $n \mapsto x_n$ *is an internal function (i.e. its graph is an internal set) and* $x_n \simeq 0$ *for all limited* n*, then there exists an unlimited* m *such that* $x_n \simeq 0$ *for all* $n \leq m$.

6.2. **Statements.** We are going to be interested in Theorems/Lemmas/Propositions (TLPs) that have statements of the form $(\forall x \in X)[\mathbb{A}(x) \Rightarrow \mathbb{B}(x)]$ and $(\exists x \in X)[\mathbb{A}(x)]$. The set X will belong either to $V(S)$ or to $V(^*S)$, and statements $\mathbb{A}(\cdot)$ can be identified with sets $A = \{x \in X : \mathbb{A}(x)\}$, this being a set in $V(S)$ or $V(^*S)$. This means that the first kind of TLP is the statement $A \subset B$, and the second kind of TLP is the statement $X \cap A \neq \emptyset$.

The **transfer principle** has a deceptively simple formulation: $A \subset B$ in $V(S)$ iff $^*A \subset {}^*B$ in $V(^*S)$; and $X \cap A \neq \emptyset$ in $V(S)$ iff $^*X \cap {}^*A \neq \emptyset$ in $V(^*S)$.