Draft, December 2013

CONTINUOUS TIME GAME THEORY: AN INFINITESIMAL APPROACH

MAXWELL B. STINCHCOMBE

Contents

1. Introduction and Overview

At an intuitive level, the infinitesimals are the df, dx, and dt's from calculus, smaller in absolute value than any "real" number, but non-zero, written $df \simeq 0$, $dx \simeq 0$, $dt \simeq 0$. Using infinitesimals allows us to define a function $f : \mathbb{R} \to \mathbb{R}$ being continuous at x if $f(x + dx) - f(x)$ is infinitesimal for any infinitesimal dx. It also allows us to define f having derivative r at a point x if $r - \frac{f(x+dx)-f(x)}{dx}$ is infinitesimal. Interest here centers on their uses in game theory models. The main themes include equilibrium refinement, control theory, differential games, continuous time games with and without diffuse monitoring, and large population games.

For this introduction, suspend disbelief for just a bit, and suppose that there exist infinitesimals, non-zero numbers smaller, in absolute value, than all of the real numbers you are used to. The next section, §2, will "construct" infinitesimals and develop the first set of tools to manipulate and use them. It is enough of a formal development that we can go a good ways further without violating any proprieties. The first set of tools is not enough to sensibly do all that we will want to do, and Section §5 will develop everything else we need.

1.1. **Equilibrium Refinement via Infinitesimals.**

A game Γ describes a strategic interaction, and $Eq(\Gamma)$ denotes the set of equilibria for Γ. If $Eq(\Gamma)$ contains many 'kinds' of equilibria, then changes in the description of the strategic situation, Γ, can have many 'kinds' of effects. As much of economic analysis is based on the dependence of equilibria on aspects of the description of a strategic situation, this can be an impediment to modeling. When the strategic situation has dynamics and differential information, this problem is widespread.

Equilibrium refinement refers to the elimination of some equilibria or sets of equilibria as being "unreasonable." This chapter covers some of the major themes in equilibrium refinement from the perturbation point of view: an equilibrium, σ^*, or set of equilibria, S, is reasonable if there are equilibria of small perturbations to the game close to σ^* or close to S. The main tool is the concept of an infinitesimal perturbation, which is an idealization of a sequence of perturbations that are converging to 0.

A game Γ, is given by a set I of players, usually people, involved in a strategic situation. Each $i \in I$ has a set of possible actions, A_i. The set of all possible choices of actions is $A := \times_{i \in I} A_i$. Preferences for each $i \in I$, are given by a von Neumann-Morgenstern utility function, $u_i : A \to \mathbb{R}$. Combining, $\Gamma = (A_i, u_i)_{i \in I}$. When there are two players and each has two actions, we call the game a 2×2 game. The starting point for perturbation-based equilibrium refinement is the following 2×2 game, due to Selten.

	$l$eft	$r$ight
up	$(1, 1)$	$(0, 0)$
$d$own	$(0, 0)$	$(0, 0)$

Some conventions: here $I = \{1, 2\}$, $A_1 = \{u\text{p}, d\text{own}\} = \{u, d\}$, $A_2 = \{l\text{eft}, r\text{ight}\} = \{l, r\}$; player 1, listed first, chooses which row occurs by picking either the action "up" or the action "down;" and player 2 chooses which column by picking either the action "left" or the action "right;" each entry in the matrix is uniquely identified by the actions a_1 and a_2 of the two players, each has two numbers, (x, y), x is the utility of player 1 and y the utility of player 2 when the vector $a = (a_1, a_2)$ is chosen, equivalently, $(x, y) = (u_1(a_1, a_2), u_2(a_1, a_2))$ when (x, y) is the entry in the a_1 row and the a_2 column.

In an equilibrium, player 1 picks action up with probability $\sigma_1 \in [0, 1]$ while player 2 picks left with probability $\sigma_2 \in [0, 1]$. $(\sigma_1^*, \sigma_2^*) = (1, 1)$ is the obvious equilibrium, $(\sigma_1^*, \sigma_2^*) = (0, 0)$ is also an equilibrium. There are many ways to talk about how unreasonable the second equilibrium is. One of the most convincing comes from the observation that 1's payoff to up are 1 or 0 depending on 2's choice of $l$eft or $r$ight, where 1's corresponding payoff to $d$own are 0 and 0. So, no matter what 2 does, 1 is at least as well off playing up, and may be strictly better off. Given the symmetry of the game, the same is true for player 2.

One way to perturb player 1's strategies is to restrict 1 to play $\sigma_1 \in [\epsilon_{1,u}, 1 - \epsilon_{1,d}]$, $\epsilon_{1,u} > 0$ and $\epsilon_{1,d} > 0$. That is, player 1 must put mass at least $\epsilon_{1,u}$ on up, and at least mass $\epsilon_{1,d}$ on $d$own. The perturbation is interior because we require that the ϵ's be strictly positive, it is infinitesimal if $\epsilon_{1,u} \simeq 0$ and $\epsilon_{1,d} \simeq 0$. The infinitesimal perturbation makes a huge difference to the set of equilibria: if 1 is playing any strategy in the perturbed set, 2's payoff to playing $l$eft is at least $1\sigma_{1,u} > 0$, while 2's payoff to playing $r$ight is 0; and the analysis is directly parallel for 1's payoffs if 2 is playing any strategy in the perturbed set. Thus, in the perturbed game, the unique equilibrium is $(\sigma_1^*, \sigma_2^*) = (1 - \epsilon_{1,d}, 1 - \epsilon_{2,r})$, and $(1 - \epsilon_{1,d}, 1 - \epsilon_{2,r}) \simeq (1, 1)$.

In general, $\Delta_i = \Delta(A_i)$ denotes the set of distributions over A_i and $\Delta := \times_{i \in I} \Delta_i$ is the set of vectors of choices with the explicit understanding that any vector $(\sigma_i)_{i \in I}$ corresponds to a product distribution. We will define an vector of strategies, $\sigma^* \in \Delta$ to be a **perfect equilibrium** if there is an interior, infinitesimal perturbation, $\Delta_i(\epsilon_i)$, to the strategy sets Δ_i and an equilibrium $\sigma^\epsilon \in \times_{i \in I} \Delta_i(\epsilon_i)$ with $\sigma^\epsilon \simeq \sigma^*$. This directly parallels the sequence definition: σ^* is a **perfect equilibrium** if there is a sequence of interior perturbations to the strategy sets, $\Delta_i(\epsilon_i^n)$, with $\epsilon_i^n \to 0$, and a sequence of equilibria, $\sigma^n \in \times_{i \in I} \Delta_i(\epsilon_i^n)$, with $\sigma^n \to \sigma^*$. We replace the sequence of perturbations with infinitesimals and replace taking limits with looking for a strategy in $\times_{i \in I} \Delta_i$ with one that is infinitesimally close.

The theory begins to have real bite and deep implications when we apply it to extensive form games. Here, we will work through the implications of stability requirements in terms of "reasonable" conjectures and/or reasonable "beliefs," especially in their iterative formulations. Stability requirements are that all equilibria, or sets of equilibria, that are not refined away, be close to the equilibrium set for all version of the game in a large class of perturbed games.

1.2. **Compact and Continuous Normal Form Games.** If $\Gamma = (A_i, u_i)_{i \in I}$ is a game, each A_i is a compact metric space, and each u_i is jointly continuous, we have the simplest class of infinite action set games. Here we can replace each set A_i with a finite set that is infinitesimally close and examine equilibrium refinement from this point of view.

1.2.1. *Limit games and limit equilibria.* As upper/lower hemicontinuity, cover results from [Fudenberg and Levine, 1986].

1.2.2. *The need for exhaustiveness.* Examples from [Simon and Stinchcombe, 1995]

1.2.3. *Respecting weak dominance.* Entry example here from [Simon and Stinchcombe, 1995], dominance arguments to the fore again.

1.3. **Extensive Form Games with Infinite Action Sets.** If we have an extensive form game with infinite sets of actions, the situation is much more complicated. Briefly cover problems in [Simon and Zame, 1990], then the fixes in [Stinchcombe, 2005].

Note: Manelli and cheap talk signaling games [Manelli, 1996]; failures of upper and lower hemicontinuity.

1.4. **Control Problems and Differential Games.** If we have an infinitesimals, dt, then we have an infinite number $1/dt$. Using $1/dt$ steps of size dt, we can approximate a time interval, e.g. $[0,1]$, by what we call a **near interval**, $T = \{0, dt, 2dt, \cdots, N \cdot dt\}$ where $N \cdot dt \simeq 1$. This turns continuous time control problems into *-finite problems, and these can often be solved using moderately elementary techniques.

1.4.1. *An Easy Control Problem.* For example, consider the following simple control problems,

$$\min \int_0^1 [c_1\dot{x}^2(t) + c_2 x(t)]\, dt \text{ s.t. } x(0) = 0, x(1) \geq B, \dot{x}(t) \geq 0. \tag{1}$$

The idea is that, starting with none of a good on hand, we need to produce a total amount B of a good by time 1. There is a storage cost, c_2, per unit stored for a unit of time, and producing at a rate r, i.e. having $x'(t) = dx/dt = \dot{x} = r$, incurs costs at a rate $c_1(x'(t))^2$. The tradeoff between producing fast productino rate and storage costs leads us to believe that the solution must involve starting production at a low level at some point in the interval and increasing the rate at which we produce as we near the end of the interval.

Let us turn to the near interval formulation of the problem. We replace $[0,1]$ by a near interval T with increments dt, and to make life simpler we suppose that the increments have equal size, $dt \simeq \frac{1}{N}$ for some infinite N, that is, we pick dt to be a special kind of infinitesimal. Now $x'(t)$ is the action, a_t, chosen at t, $a_t = \frac{x(t+dt)-x(t)}{dt}$, that is, the discrete slope of the amount on stock over the interval of time between t and $t + dt$. This means that if we choose actions $a_0, a_1, \ldots, a_{N-1}$, then by any $t \in T$, $x(t) = \sum_{s<t} a_s\, ds$. The problem is replaced by

$$\min_{a_0,a_1,\ldots,a_{N-1}} \sum_t \left[c_1 a_t^2 + c_2 \sum_{s<t} a_s\, ds\right] dt \text{ s.t. } \sum_{s<T} a_s\, ds = B,\ a_t \geq 0. \tag{2}$$

This problem can be solved by consulting the Kuhn-Tucker conditions.

In general, the existence of characterization of solutions to control problems is often easier when one uses of infinitesimals in the form of a near interval: when the near interval solution is infinitesimally close to a time path of actions in $[0,1]$, that time path is a solution to the original problem; when a control problem using $[0,1]$ as the time set does not have a solution, the near interval solutions have interpretations as *Young measures.* The following example shows what is involved in things going wrong.

1.4.2. *A Less Easy Control Problem.* Consider the problem in

$$\max \int_0^1 [\dot{x}^2(t) - x^2(t)]\, dt \text{ s.t. } -1 \leq \dot{x}(t) \leq +1 \tag{3}$$

where the maximum is taken over piecewise continuous functions $t \mapsto \dot{x}(t)$. The first term in the integrand tells us that we want to be moving as fast as possible, that is, we want $\dot{x}^2(t)$ large. The second term tells us that we want to minimize our displacement, that is, we want $x^2(t)$ small. These have a rather contradictory feel to them. Let us examine just how contradictory.

Divide the interval $[0,1]$ into N equally sized sub-intervals, $N \in \mathbb{N}$, and consider the path that over each interval $[\frac{k}{N}, \frac{k+1}{N}]$ has $\dot{x} = +1$ for the first half and $\dot{x} = -1$ for the second half. This means that $x(t)$ goes up at slope $+1$ over the first half of each interval and down with slope -1 over the second half of each interval. An $N \uparrow$, the value to this path in (82) converges up to 1. However, the value 1 cannot be achieved by any path — *that* requires that $\dot{x}(t)$ alway be either ± 1 and $x(t)$ always be 0, contradictory requirements.

Replacing $[0,1]$ with a near interval $T = \{0, \frac{1}{N}, \ldots, \frac{N-1}{N}, 1\}$ with N unlimited and even. Reformulate (3) as

$$\max_{a_0, a_1, \ldots, a_{N-1}} \sum_t \left[a_t^2 - \left(\sum_{s<t} a_s\, ds\right)^2\right] dt \text{ s.t. } -1 \leq a_t \leq +1. \tag{4}$$

Notice that there is a pair of multipliers for each $t \in T$, one for the constraint $-1 \leq a_t$ and one for the constraint $a_t \leq +1$. Only one of each of these constraints can be binding at any point in time. Often the time path of the multipliers is very informative about when and where constraints are most strongly pinching the solution. Here, there is so much symmetry that the pattern of the multipliers looks like the pattern we will see in the solutions and has no further information.

One of the two solutions to this is $a^*_{k/N} = +1$ for the even k and $a^*_{k/N} = -1$ for the odd k (the other solution reverses the signs). This gives a utility of $1 - \frac{1}{2}dt^2 \simeq 1$. We see a continuation of the pattern of approach the supremum value of 1 — since $dt = \frac{1}{N}$, larger N yields smaller dt, yielding a higher value. Thus, the near interval formulation has a solution, it gives a value $\simeq 1$.

The optimal path $t \mapsto a^*_t$ is *not* infinitesimally close to anything continuous as it move up or down by 2 between each t and $t + dt$. This is a phenomenom known as **chattering**. Not only is there not any continuous function that behaves like this, there is not any *measurable* function.

To see why, look up Lebesgue's density theorem, it tells us that for any measurable $A \subset [0,1]$, there is an $A' \subset A$ such $Unif(A \setminus A') = 0$ and for each $x \in A'$,

$$\lim_{\epsilon \downarrow 0} \frac{Unif(A \cap (x-\epsilon, x+\epsilon))}{2\epsilon} = 1. \tag{5}$$

If $A = [a,b]$, then $A' = (a,b)$, we just get rid of the end-points. The amazing part of Lebesgue's result is that this simple intuition of getting rid of end points works for all measurable sets. In particular, this means that for each $x \in A'$, the derivative of the function $H(x) = Unif(A \cap [0,x])$ is equal to 1. Applying Lebesgue's density theorem to $B := A^c$, for $x \in B'$, the derivative of $H(x)$ is equal to 0. Since $Unif(B) + Unif(B') =$

1, this means that the derivative of H is, for Lebesgue almost every $x \in [0, 1]$, either equal to 0 or equal to 1.

Now, if there is a measurable function representing $t \mapsto a^*_t$, then we can partition $[0, 1]$ into a set A on which $\dot{x}(t) = +1$ and $B := A^c$ on which $\dot{x}(t) = -1$. However, for every non-infinitesimal $x \in [0, 1]$, the proportion of the $t \in T$ with $t < x$ and $a^*_t = +1$ is, up to an infinitesimal, equal to $\frac{1}{2}$. This means that for our measurable function we would have to have $Unif(A \cap [0, x]) = Unif(B \cap [0, x]) = \frac{1}{2}x$ for each $x \in (0, 1]$. Lebesgue's density theorem tells us that this cannot happen, we must have the derivative equal to 1 or 0 almost everywhere rather than equal to $\frac{1}{2}$ everywhere.

1.4.3. *Differential Games.* Control theory is single person optimization. Equilibrium in game theory involves many people simultaneously optimizing given the others' optimizing behavior. When two or more players are simultaneously picking actions a_t, and this affects the state of the system through differential equations, we have a differential game. Differential games have a long history, but when the actions affect the state of the system through stochastic differential equations, we are in a relatively new class of games. The entry to such games passes through continuous time stochastic processes.

1.5. **Continuous Time Stochastic Processes and Diffuse Monitoring.** We cover a Brownian monitoring model and a Poisson monitoring model. Both of the continuous time processes involve are Levy processes, and studying these with near intervals is a tremendous simplification that leaves us, often enough, with the same diffuse model of monitoring others actions.

1.5.1. *Brownian monitoring.* Again let $T = \{0, \frac{1}{N}, \frac{2}{N}, \ldots, \frac{N-1}{N}, 1\}$ with N an infinite integer. Consider the probability space $\Omega = \{-1, +1\}^T$ and define P so that the canonical projection mappings $\text{proj}_t(\omega) := \omega_t$ are an i.i.d. collection with $P(\omega_t = -1) = P(\omega_t = +1) = \frac{1}{2}$. Equivalently, let P be the uniform distribution on Ω. From this, define $X(t, \omega)$ as follows: $X(0, \omega) \equiv 0$, $X(1, \omega) = \frac{1}{\sqrt{n}}\omega_1$, $X(2, \omega) = \frac{1}{\sqrt{N}}(\omega_1 + \omega_2)$, $\ldots$, $X(\frac{k}{N}, \omega) = \frac{1}{\sqrt{N}} \sum_{i=1}^{k} \omega_i$. This is a random walk model that moves through time in step sizes $dt := \frac{1}{N}$, and moves up and down $\pm\sqrt{1/N}$.

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 $\text{Var}(X(\frac{k}{N}, \cdot)) \simeq r$. The oldest (deMoivre) arguments for the central limit theorem tells us 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. This is infinitesimally close to a Brownian motion. By changing the probabilities of $\pm\sqrt{1/N}$ by the appropriate infinitesimal, the Brownian motion gains a drift. If the drift depends on the action of one player, the process provides a diffuse and noisy signal of that action, one in which evidence becomes stronger and stronger over time.

1.5.2. *Poisson monitoring.* In a similar vein, continuing with $T = \{0, \frac{1}{N}, \frac{2}{N}, \ldots, \frac{N-1}{N}, 1\}$, 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 $dt := \frac{1}{N}$ is the infinitesimal size of the incremental steps in the time set, and λ is limited and strictly positive. Define $Y(0,\omega') \equiv 0$, $Y(\frac{1}{N},\omega') = \omega'_1$, $Y(\frac{k}{N},\omega') = \sum_{i\leq k} \omega'_i$.

If $r \in (0,1]$ and $\frac{k}{N} \simeq 0$, $Y(\frac{k}{N},\cdot)$ is infinitely close to having a Poisson(λr) distribution (by the binomial approximation to Poisson distributions). 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. This is infinitesimally close to a Poisson process. By changing the infinitesimal probability λdt to $\lambda' dt$, we change the arrival rate of the Poisson process. If the arrival rate depends on the action of one player, the process provides a diffuse and noisy signal of that action, one in which evidence becomes stronger and stronger over time.

Exercises: the minimum of two negative exponentials is a negative exponential with half the mean and cetera; distributions of Erlangs can be had from this; hypoexponentials and Coxian distributions from intensity matrixes; use in queuing theory.

1.5.3. *Levy processes and monitoring.*

1.6. **Continuous Time Game Theory with Sharp Monitoring.** In the monitoring of others' actions is not diffuse but instantaneous, we have a very different class of games. Consider a game in which players change between different actions at any time $t \in [0,1]$, and receive utility $\int_0^1 u_i(t,a(t))\,dt$ where $a(t)$ is the vector of actions picked at time t. There are difficulities defining strategies if the players can respond instantly to other players' changes of action. These definitional difficulties disappear if we replace $[0,1]$ with a near interval $T = \{0, \frac{1}{N}, \frac{2}{N}, \ldots, \frac{N-1}{N}, 1\}$. With this formality in hand, we can fruitfully study continuous time folk theorems, adoption games, price competition.

The situation is much as we will see in the analysis of control problems: when the near interval equilibrium path is infinitesimally close to a time path of actions in $[0,1]$, that time path is an equilibrium to the original problem; when a game using $[0,1]$ as the time set does not have a solution, the near interval equilibrium will have interpretations as *Young measures.*

1.7. **Large Population Games.** The unit interval, $[0,1]$, has served not only as a time set, but as a model of an infinite population of players. The attraction of this model is that each player is atomistic, infinitely small, and their actions are easily modeled as not having any effect on anyone but themselves. However, and this is the basis for how useful the model is, any positive fraction of the population does have an effect on others.

Such models are often motivated as idealizations of very very large finite sets of players. However, there are several types of equilibrium results using $[0,1]$ as the set of players that are not limits of equilibrium results using sequences of equilibrium in models with larger and larger finite sets of players. Replacing $[0,1]$ by $\{0, \frac{1}{N}, \frac{2}{N}, \ldots, \frac{N-1}{N}, 1\}$ means that the limit results hold automatically, providing a better basis for understanding the large population models.

2. Constructing Infinitesimals and the Tools to Use Them

Consider the following four sequences, three of them with the same limit:

$$r = (r, r, r, r, \ldots) \tag{6}$$
$$\epsilon = (1, 1/2, 1/3, 1/4, 1/5, \ldots) \tag{7}$$
$$\epsilon^2 = (1, 1/4, 1/9, 1/16, 1/25, \ldots) \tag{8}$$
$$0 = (0, 0, 0, 0, 0, \ldots) \tag{9}$$

For any $r \in \mathbb{R}$, $r > 0$, for all but a small set of integers n, we have

$$0 < \frac{1}{n^2} < \frac{1}{n} < r. \tag{10}$$

One development of infinitesimals adds new numbers to $\mathbb{R}$, denoting the new enlarged set as ${}^*\mathbb{R}$, by identifying numbers as equivalence classes of sequences in such a fashion that none of the four sequences are equivalent — they are unequal for all but finitely many indices, therefore they are not equal/equivalent.

Defining "$^* <$" the same way, we say that $0^* < \epsilon^{2*} < \epsilon^* < r$ because $\{n \in \mathbb{N} : 0 < \frac{1}{n^2} < \frac{1}{n} < r\}$ has at most a finite complement.[1] These inequalities are an indication that we are keeping track of what is happening on the way to the limit. The small, non-zero sequences are the non-zero "infinitesimals," written $\epsilon \simeq 0$ and $\epsilon^2 \simeq 0$, because their absolute value is smaller than any real number $r > 0$, yet they are strictly non-zero, in this case strictly positive.

If F_n is a sequence of finite sets, e.g. $F_1 = \{0, \frac{1}{2^1}, 1\}$, $F_2 = \{0, \frac{1}{2^2}, \frac{2}{2^2}, \frac{3}{2^2}, 1\}$, ..., $F_n = \{\frac{k}{2^n} : 0 \leq k \leq 2^n\}$, we get a ***-finite** set by considering the equivalence class of this sequence. This is one example of the near intervals we mentioned above. What is crucial here is that we keep track of how we get to the limit from inside the set. In a bit more detail, $x_n \neq x'_n \in F_n$, $x_n \to x$, and $x'_n \to x$, then x belongs to the limit set for the sequence of sets, and we got to that limit two different ways. What *-finite sets do is to keep track of the ways in which you converge to points in the limit set.

2.1. A Purely Finitely Additive Point Mass.

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 $K + 1$.

[1] For notational simplicity, we will drop the "*" in from of "<."

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 $(A \cap B)$.

2.2. **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 2.2.1. *${}^*[0,1]$ 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 2.2.2. *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 2.2.3. *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 2.2.4. *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.*

2.3. **Normal Form Equilibrium Refinement.** Note: expand this section with GT-Notes, varieties of dominance, iterated procedures.

2.3.1. *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 2.3.1. *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 $Eq(\Gamma)$.*

2.3.2. *Perfection.* 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 2.3.2. *For $\epsilon \in {}^*\mathbb{R}_{++}$, $\mu \in \times_{i \in I} {}^*\Delta_i^\circ$ is ϵ-**perfect** if*

$$(\forall i \in I)(\forall b_i \in A_i)[\,[\max_{a_i \in A_i} u_i(\mu \setminus a_i) > u_i(\mu \setminus b_i)] \Rightarrow [\mu_i(b_i) < \epsilon]\,]. \qquad (11)$$

*$\mu^* \in \times_{i \in I} \Delta(A_i)$ 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 2.3.3. *$\mu^* \in \times_{i \in I} \Delta(A_i)$ is **strictly perfect** if for all $\mu \in \times_{i \in I} {}^*\Delta_i^\circ$,*

$$[\,[\mu \simeq \mu^*] \Rightarrow (\forall i \in I)[\mu_i(Br_i(\mu) \simeq 1]\,]. \qquad (12)$$

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

2.3.3. *Properness.*

Definition 2.3.4. *For $\epsilon \in {}^*\mathbb{R}_{++}$, $\mu \in \times_{i \in I} {}^*\Delta_i^\circ$ is ϵ-**proper** if*

$$(\forall i \in I)(\forall a_i, b_i \in A_i)[\,[u_i(\mu \setminus a_i) > u_i(\mu \setminus b_i)] \Rightarrow [\mu_i(b_i) < \epsilon \cdot \mu_i(a_i)]\,]. \qquad (13)$$

*$\mu^* \in \times_{i \in I} \Delta(A_i)$ 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 $Pro(\Gamma)$.*

2.3.4. *p-Stability.*

Definition 2.3.5. *A closed connected set $S \subset Eq(\Gamma)$ is **robust to perturbations** if*

$$(\forall \mu \in \times_{i \in I} {}^*\Delta_i^\circ)[\,[d_H(\mu, {}^*S) \simeq 0] \Rightarrow (\forall i \in I)[\mu_i(Br_i(\mu)) \simeq 1]\,]. \qquad (14)$$

*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.*

2.4. **The Basic Results.**

A. We will 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 $S \cap Pro(\Gamma) \neq \emptyset$.

B. 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.]

Example 2.4.1. *The perfect equilibria for the following game 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	$(-2,-1)$	$(-2,0)$	$(-2,-2)$

Example 2.4.2. *The following game has no strictly perfect equilibrium, but its p-stable set of equilibria is nice.*

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

2.5. **Compact and Continuous Games.**

3. Extensive Form Games

Agent normal form, implications.

3.1. **Decision Theory with Full Support Probabilities.**

Readings for this section are [Blume et al., 1991a] and [Blume et al., 1991b].

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.

3.2. **Dynamic Decisions, 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 $E_1, \ldots, E_K$ 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, $k \in \{1, \ldots, K\}$ 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,

$$(\forall A \in \mathcal{F}) \left[P(A) = \sum_k P(E_k) \frac{P(A \cap E_k)}{P(E_k)} = \sum_k P(E_k) P(A|E_k) \right]. \tag{15}$$

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 $P_k = P(E_k)$.

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

$$P: \quad \max_{a_1,\ldots,a_K \in A} \int 1_{E_k}(\omega) u(a_k, \omega)\, dP(\omega). \tag{16}$$

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

$$P_k: \quad \max_{a \in A} \int u(a, \omega)\, dP(\omega|E_k). \tag{17}$$

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 3.3.1 (Bridge Crossing). *$(a_1^*, \ldots, a_K^*)$ 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.

3.4. **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 3.4.1. *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.

3.5. **Extensive Form Equilibrium Refinement.**

3.5.1. *Perfect, proper, and stable equilibrium outcomes.*

3.5.2. *Iterative dominance arguments.*

4. Continuous Time Control and Games

4.1. **Control Theory.** The need for a theory of integration.

4.1.1. *When it Works Wonderfully.*

4.1.2. *When it Works Less Well.* On the importance of existence theorems for the interpretation of necessary conditions, we have Perron's paradox:

"If N is the largest positive integer, then $N = 1$. To see why, suppose that N is the largest positive integer but $N > 1$. Then $N^2 > N$, implying that N was not the largest positive integer."

To put it another way, a necessary condition for N being the largest positive integer is that $N = 1$. However, since there is no largest positive integer, the necessary condition is non-sense.

Formally, conditioning on N being the largest positive integer, i.e. conditioning on having an element in the null set, we can derive all kinds of things.

Fillipov's Theorem: For the basic version and examples of this, see p. 119 of Liberzon's "Calc of Variations etc." (which is on the iPad as "cvoc"), [Liberzon, 2012],

A grown-up version is in the following and its prequel, [McShane and Warfield, 1969]

4.2. **Games with Instantaneous Monitoring.**

4.2.1. $(\mathbb{R}, \leq)$ *is Totally Ordered but Not Well-Ordered.*

Definition 4.2.1. *A set X is* ***totally ordered*** *by a relation $\preceq \subset X \times X$ if for every $x, y, z \in E$,*
(i) $[x \preceq y] \wedge [y \preceq x] \Rightarrow [x = y]$,
(ii) $[x \preceq y] \wedge [y \preceq z] \Rightarrow [x \preceq z]$, *and*
(iii) $[x \preceq y] \vee [y \preceq x]$.

A chain of sets is totally ordered by $\subseteq$. Any time set $X \subset \mathbb{R}$ is totally ordered by $\leq$. For the purposes of game theory, we need to go from what the agents choose to do now to what happens next.

Definition 4.2.2. *A totally ordered set $(X, \preceq)$ is* ***well-ordered*** *if every non-empty $S \subset X$ contains a least element.*

A chain of subsets of the integers is totally ordered by $\subseteq$. Any countable time set $X \subset \mathbb{R}$ that has no accumulation points from the right is totally ordered by $\leq$.[2]

Lemma 4.2.3. *If there exists a $t \in X \subset \mathbb{R}$ and a sequence x_n in X with $x_n \downarrow t$, then there is no least element in X that is greater than t.*

Proof. If there exists $t' \in X$, $t < t'$, but $X \cap (t, t') = \emptyset$, then we have a contradiction to $x_n \downarrow t$. □

We now turn to some of the problems this causes for specifying strategies in continuous time games.

4.2.2. *Implications.*
In discrete time, and in any near interval, for each time t, players look at what they know of the history so far and choose what they will do at the "next time" they have a chance to move. There is no such "next time" in continuous time modeled as $[0, \infty)$ or $[0, 1]$ or $[0, 1)$. This already causes problems for specifying how strategies map to outcomes in single-agent 'games,' and causes even worse problems for games with two or more agents.

A strategy should specify, for each t and each history on the interval $[0, t)$, what action should be chosen at t. A sensible definition of the outcome associated with a choice of strategies is one that agrees, at each point in time, t, with what the strategy calls for at t.

Example 4.2.1. *Suppose that $I = \{1\}$, the action set is $A_1 = [0, 1]$, and at every $t \in [0, \infty)$, player 1 plays the strategy $\sigma_1(h, 0) = 0$ and $\sigma_1(h, t) = \sup\{h(s) : 0 \leq s < t\}$ for $t > 0$. For any $\tau > 0$, any history of the form $h_\tau(t) = 0$ for $t \leq \tau$, $h_\tau(t) = r$ for $t > \tau$ and $r \neq 0$ has the property that for all $t \in [0, \infty)$, $\sigma_1(h_\tau, t) = h_\tau(t)$.*

Intuitively, the strategy calls for the player to start by playing 0 and to continue playing 0 so long as 0 is all that has been played in the past. It is a well-formed strategy in the sense that it maps all histories and all times t into a definite choice of action. The problem is that there is no first time past τ at which the strategy is "mistaken."

[2] $t \in X$ is an accumulation point from the right for X if $(\exists x_n in X)[x_n \downarrow t]$.

There are even worse problems when there are two or more players. Suppose that at any $t \in [0, \infty)$, players $i \in I = \{1, 2\}$ are taking an action in $A_i = \{0, 1\}$. This means that a history is a function $h : [0, \infty) \to A$, $A = \times_{i \in I} A_i$. We can represent each history h as a pair of histories giving the actions of the two players, $h = (h_1, h_2)$ where $t \mapsto h_1(t) \in A_1$ and $t \mapsto h_2(t) \in A_2$. For any history $h = (h_1, h_2)$ and $t > 0$, let $h_{|t-} : [0, t) \to A$ denote the truncation of h before t, that is, $h_{|t-}(s) := h(s)$ for $s < t$.

A strategy for $i \in I$ specifies $\sigma_i(h, t) \in A_i$ with the restriction that $[h'_{|t-} = h_{|t-}] \Rightarrow [\sigma_i(h, t) = \sigma_i(h', t)$. A set $X \subset (0, \infty)$ is **dense-in-itself from the left** if for each $x \in X$ and $\epsilon > 0$, $X \cap (x - \epsilon, x) \neq \emptyset$: $\mathbb{Q} \cap (0, \infty)$ is dense-in-itself from the left; if $X \cap (0, \infty)$ has full Lebesgue measure, then it is dense-in-itself from the left. The following is from [Stinchcombe, 1992].

Example 4.2.2. *The following strategies capture the following mis-matching ideas: if player* 2 *has been playing* $a_2 = 0$ *recently,* 1 *wants to mis-match, that is to play* $a_1 = 1$*, otherwise they want to mis-match by playing* $a_1 = 0$*; if player* 1 *has been playing* $a_1 = 1$ *recently, player* 2 *wants to match by playing* $a_2 = 1$*, otherwise wants to mis-match by playing* $a_2 = 0$*. More precisely, let* $\sigma_i(h, 0) = 0$ *for* $i \in I$*, and for* $t > 0$*, let*

$$\sigma_1(h, t) = \begin{cases} 1 & \textit{if } \limsup_{s \uparrow t} h_2(s) = 0 \\ 0 & \textit{else}, \end{cases} \qquad \sigma_2(h, t) = \begin{cases} 1 & \textit{if } \liminf_{s \uparrow t} h_1(s) = 1 \\ 0 & \textit{else}. \end{cases} \tag{18}$$

Being explicit about the "else" cases is useful: $\neg \limsup_{s \uparrow t} h_2(s) = 0$ *iff* $\limsup_{s \uparrow t} h_2(s) = 1$*;* $\neg \liminf_{s \uparrow t} h_1(s) = 1$ *iff* $\liminf_{s \uparrow t} h_1(s) = 0$.

It is one thing to say that there is no function agreeing with $\sigma(h, t)$ at all t or at most t. The next result says much more, it says that for any possible history h, $\sigma(h, t)$ disagrees with $h(t)$ at all strictly positive points in the time set.
Claim: If $h : X \to A$ and X is dense-in-itself from the left, then for all $t \in X \cap (0, \infty)$, $h(t) \neq \sigma(h, t)$.

Proof. At any $t \in X \cap (0, \infty)$, either $\sigma_1(h, t) = 1$, i.e. $\limsup_{s \uparrow t} h_2(s) = 0$, or $\sigma_1(h, t) = 0$, i.e. $\limsup_{s \uparrow t} h_2(s) = 1$.
Case 0, $\limsup_{s \uparrow t} h_2(s) = 0$: This can only happen if $(\exists \epsilon > 0)(\forall s \in X \cap (t - \epsilon, t))[h_2(s) = 0]$. To be consistent with σ_1, this means that $(\forall \tau \in X \cap (t - \epsilon, t))[h_1(\tau) = 1]$. To be consistent with σ_2, this in turn means that $(\forall \tau \in X \cap (t - \epsilon, t))[h_2(\tau) = 1]$, which contradicts $\limsup_{s \uparrow t} h_2(s) = 0$.
Case 1, $\limsup_{s \uparrow t} h_2(s) = 1$: This can only happen if $(\forall \epsilon > 0)(\exists \tau \in X \cap (t - \epsilon, t))[h_2(\tau) = 1]$. For $h_2(\tau) = 1$ to be consistent with σ_2, there must exists $\delta > 0$ such that $h_1(s) = 1$ for all $s \in X \cap (\tau - \delta, \tau)$. Consistency with σ_1 implies that $h_2(s) = 0$ for $s \in X \cap (\tau - \delta, \tau)$, which contradicts σ_1 and $h_1(s) = 1$ on this interval. □

4.3. **Games Played on a Near Interval.** The paper [Simon and Stinchcombe, 1989] gave an expanded history space, one that allows for infinitely fast reactions, and conditions on strategies guaranteeing that playing the strategies on any near interval gave rise to histories with two properties: the standard parts belong to the expanded history space; and the outcomes are at infinitesimal distance from each other. The paper [Stinchcombe, 1992] used a more general version of the history space and gave the minimal conditions on vectors of strategies guaranteeing playability on standard time intervals. Here, we will study games with flow payoffs on near intervals, examine the set of equilibria, and when we do and do not have nearstandard equilibrium outcomes in the expanded history space.

4.3.1. *Actions, Histories, Strategies, and Outcomes.* Each $i \in I$ has a non-empty action set A_i, A denotes $\times_{i\in I} A_i$. Time is $T = \{t_k = \frac{k}{N} : k = 0, \ldots, N\}$, $N = m!$ for $m \simeq \infty$. Note that $t_0 = 0$, $t_N = 1$, $dt = \frac{1}{N} \simeq 0$, and T is a near interval. A **complete history** is a point in $\{\gamma\} \times A^T$, but we will mostly notationally suppress the γ. For $t \in T$, $t > 0$, $t-$ denotes the largest $t_k \in T$ with $t_k < t$, and $h_{|t-}$ denotes the restriction of h to the set $\{0, \ldots, t-\}$, i.e. $h_{|t-} \in A^{\{0,\ldots,t-\}}$ and $h_{|t-} = \text{proj}_{\{0,\ldots,t-\}}(h)$. We define $h_{|0-} = \gamma$.

A **pure strategy for** $i \in I$ specifies what $a_i \in A_i$ player i chooses at $t = 0$, and what they choose in response to each $h_{|t-}$. A **behavioral strategy for** $i \in I$ specifies what $\sigma_i \in \Delta(A_i)$ they pick. With the γ convention above, the set of decision nodes is $\cup_{t\in T}\{h_{|t-} : h \in A^T\}$, and behavioral strategies map decision nodes to mixtures over the A_i's.

An **outcome** is a distribution over A^T. For each decision node, $h_{|t-}$, and vector of behavioral strategies, σ, there exists a unique outcome $\mathbb{O}(h_{|t-}; \sigma)$ defined in the usual inductive way, though now the induction is over the near interval.

For each $h \in A^T$, $U_i(h) := \sum_{t\in T} u_i(t, a_t)\, dt$. We extend each U_i to strategies starting at decision nodes in the usual fashion, $U_i(h_{|t-}; \sigma) = \sum_H U_i(h)\mathbb{O}(h_{|t-}; \sigma)(h)$, that is, $U_i(h_{|t-}; \sigma)$ is the expected utility if σ is played starting at $h_{|t-}$.

A game on a near interval T is given by $\Gamma = (H, (u_i(t; \cdot))_{t\in T})_{i\in I}$ where $H \subset A^T$. The strategies must be restricted so that only outcomes in H arise. By choosing H judiciously, we can cover a number of different types of games: if H contains only histories in which each $i \in I$ starts at $a_i = 0$ and changes action only once, then strategies can only give the option to change if no change has happened before; if H contains only histories in which i's action is constant on some sub-interval of T, then i's strategies must be restricted so that the cannot move during that sub-interval; if H contains only histories in which i chooses in a set $A'_i \subset A_i$ if j has previously chosen in a set $A'_j \subset A_j$, then j's choices determine i's options; if at each $h_{|t-}$, each $i \in I$ can only choose from some non-empty $A_i(h_{|t-}) \subset A_i$, then H becomes the set of histories with $h(t) = \times_{i\in I} A_i(h_{|t-})$; more subtle restrictions on H allow for a choice at t_k to affect possible choice sets of some or all agents at t_{k+m}.

We will be, at first, mostly interested in the case that $t \mapsto u_i(t, \cdot)$ is constant, or at least, near-continuous. Then, to handle timing and pre-emption games, we will restrict to subsets of $H \subset A^T$ where the agents start move at most once, corresponding to the choice to enter a market. We call Γ a **continuously repeated game** if $H = A^T$.

Definition 4.3.1. *σ^* is an **equilibrium** if for all $i \in I$ and all strategies σ_i for i, $U_i(h_{|0-}; \sigma^*) \geq U_i(h_{|0-}; \sigma^* \backslash \sigma_i)$; it is an **infinitesimal equilibrium** if for all $i \in I$ and all strategies σ_i for i, ${}^\circ U_i(h_{|0-}; \sigma^*) \geq {}^\circ U_i(h_{|0-}; \sigma^* \backslash \sigma_i)$.*

Infinitesimal equilibria are hybrid objects: the strategy spaces belong to $V({}^*S)$; the utility function is external, it does not belong to $V({}^*S)$. A Loeb space $(\Omega, L(\mathcal{F}), L(P))$

is a hybrid objects in much the same way: the set Ω is internal, the class of sets $\mathcal{F}$ is internal, the function P is internal; for $E \in \mathcal{F}$, $L(P)(E) = {}^\circ P(E)$, and we expand the domain of $L(P)$ to the completion of the sigma-field generated by $\mathcal{F}$.

Calculations in Loeb spaces are often simplified by being able to show that some probability is infinitesimal, hence can be ignored. We have a similar situation for infinitesimal equilibria: σ^* is an infinitesimal equilibrium iff for some infinitesimal ϵ, $U_i(h_{|0-}; \sigma^*) + \epsilon \geq U_i(h_{|0-}; \sigma^* \backslash \sigma_i)$.

Definition 4.3.2. *σ^* is a **subgame perfect equilibrium (sgpe)** if for all $i \in I$, all decision nodes, $h_{|t-}$, and and all strategies σ_i for i, $U_i(h_{|t-}; \sigma^*) \geq U_i(h_{|t-}; \sigma^* \backslash \sigma_i)$; it is an **infinitesimal subgame perfect equilibrium (isgpe)** if for all $i \in I$, all decision nodes $h_{|t-}$, and all strategies σ_i for i, ${}^\circ U_i(h_{|t-}; \sigma^*) \geq {}^\circ U_i(h_{|t-}; \sigma^* \backslash \sigma_i)$.*

To check that a strategy is a sgpe, we need only check that there are no profitable one-period deviations — a complicated deviation can not gain utility if at every step it is losing, or not gaining, utility. To check that a strategy is an isgpe, it seems that we may need to do more — summing an infinite number of infinitesimals utility gains may give a non-infinitesimal utility gain. However there is a sufficient finite deviation condition: an internal deviation from an isgpe yields an internal sequence of utility gains, $(r_t)_{t \in T}$; to be an isgpe, we must have $\sum_{t \in T} r_t \, dt \simeq 0$; if, for example, $\frac{r_t}{dt}$ is infinitesimal except for a subset $T' \subset T$ with $|T'|/|T| \simeq 0$, then $\sum_{t \in T} r_t \, dt \simeq 0$; in particular, finiteness of T' is sufficient, and as we will see, this is often enough for our purposes.

4.3.2. *Safety in Continuously Repeated Games.* For a continuously repeated game, there are three "safe" utility levels that one might imagine i being able to guarantee $i$self at time t, and their value integrated over the course of the game,

$$\underline{v}_{i,t}^{pure} = \min_{a_{-i} \in \times_{j \neq i} A_j} \left[\max_{b_i \in A_i} u_i(t; b_i, a_{-i}) \right], \; V_i^{pure} := \sum_{t \in T} v_{i,t}^{pure} \, dt, \tag{19}$$

$$\underline{v}_{i,t}^{mixed} = \min_{\sigma_{-i} \in \times_{j \neq i} \Delta(A_j)} \left[\max_{b_i \in A_i} u_i(t; b_i, \sigma_{-i}) \right], \; V_i^{mixed} := \sum_{t \in T} v_{i,t}^{mixed} \, dt \text{ and} \tag{20}$$

$$\underline{v}_{i,t}^{corr} = \min_{\mu_{-i} \in \Delta(\times_{j \neq i} A_j)} \left[\max_{b_i \in A_i} u_i(t; b_i, \mu_{-i}) \right], V_i^{corr} := \sum_{t \in T} v_{i,t}^{corr} \, dt. \tag{21}$$

Since $\times_{j \neq i} A_j \subset \times_{j \neq i} \Delta(A_j) \subset \Delta(\times_{j \neq i} A_j)$, $\underline{v}_{i,t}^{pure} \geq \underline{v}_{i,t}^{mixed} \geq \underline{v}_{i,t}^{corr}$.

Once we understand what these safety levels are about, it is easy to give games where the inequalities are strict: the first safety level corresponds of the worst that dolts who do not understand randomization can do to i; the second corresponds of the worst that enemies who do understand independent randomization can do to i; the third corresponds of the worst that fiends who completely understand randomization can do to i. The three $\underline{v}_i$'s are called "safety levels." Here is one of the reasons.

Lemma 4.3.3. *If σ is an equilibrium of a continuously repeated game, then $U_i(\sigma) \geq V_i^{mixed}$, if it is an infinitesimal equilibrium, then ${}^\circ U_i(\sigma) \geq {}^\circ V^{mixed}$.*

This lemma is ridiculously easy to prove once you see how.

Proof. For any strategy σ_{-i} for the other players, consider the strategy that at time t myopically best responds to $\sigma_{-i}(h_{|t-})$. For each $t \in T$, the associated utility is at least $v_{i,t}^{mixed}$. □

4.3.3. *Some Examples.* An easy one to analyze is the *Prisoners' Dilemna*: $I = \{1, 2\}$, $A_1 = A_2 = \{Sil, Sq\}$, $u(t, \cdot)$ independent of t and given by

	Sil	Sq
Sil	$(3, 3)$	$(-1, 4)$
Sq	$(4, -1)$	$(0, 0)$

Here $V^{pure} = V^{mixed} = V^{corr} = (0, 0)$. If σ is an equilibrium, then $U(\sigma) = (0, 0)$. If $\epsilon \simeq 0$ is larger than dt, then the Nash reversion strategies are an ϵ-sgpe with $U(\sigma) = (3, 3)$. Further, $\text{st}\,\{U(\sigma) : \sigma \text{ is a } \epsilon - sgpe\} = \{v \in \text{con}\,(u(A)) : v \geq (0, 0)\}$.

Rational Pigs is approximately as easy to analyze: $I = \{1, 2\}$, $A_1 = A_2 = \{P, W\}$, $u(t, \cdot)$ independent of t and given by

	P	W
P	$(-1, 4)$	$(-1, 5)$
W	$(2, 2)$	$(0, 0)$

Again, $V^{pure} = V^{mixed} = V^{corr} = (0, 0)$ and $\text{st}\,\{U(\sigma) : \sigma \text{ is an } \epsilon - sgpe\} = \{v \in \text{con}\,(u(A)) : v \geq (0, 0)\}$.

Matching Pennies: $u(t; \cdot)$ is given by

	H	T
H	$(+1, -1)$	$(-1, +1)$
T	$(-1, +1)$	$(+1, -1)$

Here there is a unique equilibrium outcome, involving $(\frac{1}{2}, \frac{1}{2})$ randomization in each period. The associated equilibrium outcome has no standard part.

An interpretational issue: For Matching Pennies, all ϵ-sgpe equilibria, $\epsilon \simeq 0$ involve chattering, as do many of the equilibria discussed in the previous two examples. Chattering paths on a near interval have no standard counterpart. There are a number of ways to force the outcomes to have standard paths: the easiest follows [Simon and Stinchcombe, 1989], restrict to $H \subset A^T$ with only finitely many actions for each player; in a similar vein, make changes of actions costly enough that it will never be optimal to engage in more than finitely many; more subtly, follow [Bergin and Macleod, 1993] or [Perry and Reny, 1993] and require that players have some form of inertia in their choices.

The general result for continuously repeated games is that if $t \mapsto u(t; \cdot)$ is near continuous, then $\text{st}\,\{U(\sigma) : \sigma \text{ is an } \epsilon - sgpe\}$ is the intersection of the convex hull of $U(A^T)$ and $\{v : v \geq V^{mixed}\}$. The proof uses strategies a bit more complicated than those in the examples so far, and essentially the same proof applies if $t \mapsto u(t; \cdot)$ is a lifting of a measurable function.

4.3.4. *Revisiting Cournot vs. Bertrand.* Simple two-firm Cournot quantity competition models have equilibrium prices above marginal costs but below marginal revenue. Simple Bertrand competition models with identical firms have a unique equilibrium with $p^* = C'(q(p^*)/2)$ where $C(\cdot)$ is the cost function for the firms and $p \mapsto q(p)$ is the demand curve. Bertrand used this argument to assail Cournot's model of competition. In continuous time, they both have, as an equilibrium, that the two firms split monopoly profits.

To make things really stark, let us suppose that the technology is such that both firms can supply the whole market without disadvantaging themselves, $C(q) = c \cdot q$. Consider the near interval game with $u_i(t; p_i, p_j)$ given as follows,

$$u_i(t; p_i, p_j) = \begin{cases} (p_i - c) \cdot q(p_i) & \text{if } p_i < p_j, \\ \frac{1}{2}(p_i - c) \cdot q(p_i) & \text{if } p_i = p_j, \\ 0 & \text{if } p_i > p_j. \end{cases} \tag{22}$$

Here $A_i = {}^*[0, \overline{p}]$, $\overline{p}$ being near standard, or A_i is a near interval of prices. In either case, let p_{Mon} be the monopoly price and π_{Mon} the associated industry profits.
Claim: Both firms charging the monopoly price for each $t \in T$ is an ϵ-sgpe for $\epsilon \simeq 0$.

Proof. The grim-trigger strategies "start by playing p_{Mon}, continue to play p_{Mon} as long as that is all that has been played in the past, else play $p = c$" are ϵ-sgpe provided $\epsilon > \pi_{Mon} dt$. □

Problem 4.3.1. *An ϵ-sgpe of the continuously repeated Cournot model, $\epsilon \simeq 0$, yields the same "split monopoly profits" payoffs.*

4.3.5. *Preemption Games.* Here we examine [Fudenberg and Tirole, 1985].

Let $I = \{1, 2\}$, $A = \{\text{Out}, \text{In}\}$, and let $H \subset A^T$ be the set of time paths where players change their action at most once, and if there is a change, it is from Out to In. F&T interpret this switch of actions as the adoption of a new technology. Entering early and being alone will be good for the firm because they enjoy monopoly profits, however, entering early will incur a larger cost than entering late.

Payoffs have two parts, here divided into a flow part and a lump part, one can usually convert a lump cost/benefit into its value-equivalent flow.

Flow payoffs : $\pi_O(0)$ is the net cash flow of firm i when 0 firms have entered and i is still out; $\pi_O(1)$ is the net cash flow of firm i when the other firm has entered and i is still out; $\pi_I(1)$ is the net cash flow if i has entered and is the only firm in the industry; and $\pi_I(2)$ is the the net cash flow if both have entered.

Lump costs : The cost of entering at t is $c(t)$, and it falls over time as the technology becomes more mature.

$V_1(t_1, t_2)$ will denote 1's flow profits if 1 adopts at t_1 and 2 at t_2.

i. If both firms adopt at t, then $V_1(t, t) = V_2(t, t) = \int_0^t \pi_O(0) e^{-rt}\, dt + \int_t^\infty \pi_I(2) e^{-rt}\, dt$.

ii.

If both firms adopt at t, then

Let $L(t)$ denote the payoffs to being the Lead firm to enter if entry happens at t and the other firm best responds, $F(t)$ the payoffs to being the Follower firm if the firm enters at t, and $M(t)$ the payoffs if the firms enter si$M$ultaneously at t. These functions are continuous, and F&T make assumptions on the flow payoffs that imply that no firm will adopt at $t = 0$, and that there exists $0 < T_1 < T_1^* < T_2^* < \widehat{T}_2 < \infty$ such that

- T_1^* is a firms favorite time to adopt if the other firm follows and plays a best response;
- T_2^* is the follower's best response adoption time to any $t < T_2^*$;
- $L > F > M$ on (T_1^*, T_2^*);
- $L < F$ on $(0, T_1^*)$;
- T_1^* maximizes $L(t)$; and
- $\widehat{T}_2$ maximizes $M(t)$.

The definition of (T_1^*, T_2^*) make them the unique equilibrium of a precommitment game, that is, a game in which the two firms can commit to an entry time.

There are four results about the dynamic game.

1. If $L(T_1^*) > M(\widehat{T}_2)$, the unique equilibrium outcome has, with probability $\frac{1}{2}$, firm one adopting at T_1 and firm two adopting at T_2^*, and with probability $\frac{1}{2}$ the reverse.
2. If $L(T_1^*) < M(\widehat{T}_2)$, then, defining S as the solution to $M(s) = L(T_1^*)$, there are two classes of equilibria: the previous (T_1, T_2^*) equilibria; pure strategy equilibria with joint adoption at every $t \in [S, \widehat{T}_2]$.
3. If the pure strategy equilibria exist, they Pareto dominate the precommitment equilibrium, and the precommitment equilibria dominate the (T_1, T_2^*) equilibrium.
4. The Pareto rank of the pure strategy equilibria adopting at $t \in [S, \widehat{T}_2]$ is given by $\leq$.

The second to last result has a "firms hate competition" flavor to it. Among the pure strategy equilibria, one can pick using a weak dominance argument: play the strategy respond instantly to adoption at $t \in [S, \widehat{T}_2)$, adopt at $\widehat{T}_2$ whether or not the other has adopted (and fill in subgames past $\widehat{T}_2$ sensibly).

Let us consider a t_k in the interval $(T_1, T_2^*]$ and a subgame $h_{|t_k-}$ at which no-one has yet adopted. Consider pure strategy equilibria, hard to find any. Symmetric mixed equilibria where $U(t_k)$ is the value of going on to t_{k+1} with no-one having adopted. This has the following form,

	In	Out
In	$(M(t_k), M(t_k))$	$(L(t_k), F(t_k))$
Out	$(F(t_k), L(t_k))$	$(U(t_k), U(t_k))$

If $(\gamma_k, 1-\gamma_k)$ in (In, Out) is an equilibrium, then it has payoffs $\gamma_k M(t_k) + (1-\gamma_k)L(t_k)$. This leads to $\gamma_k = \frac{L(t_k)-F(t_k)}{L(t_k)-M(t_k)}$ up to an infinitesimal ratio.

A useful second step is the following lemma, especially if you think of r as $\frac{d}{dt}(L(t) - F(t))_{|t=T_1}$.

Lemma 4.3.4. *If T is a near interval and X_k is a collection of independent random variables with $P(X_k = 1) = rt_k$ and $P(X_k = 0) = 1 - rt_k$ for a non-infinitesimal r, then for some infinitesimal $t_K \in T$, $P(\sum_{k \leq K} X_k = 0) \simeq 0$.*

4.3.6. *Wars of Attrition.* Two competitors face each other in a battle of wills, s/he who hangs on the longest wins. This can be interpreted in a variety of ways: as a contest between firms in an industry only large enough to support one of them; as a contest between animals for a valuable prize, be it a mating opportunity, a prime spot for a nest, a food source; as a republican budget strategy.

We suppose that the value to competitor i is $s_i \geq 0$, that s_i and s_j are iid with cdf F. The choice of action is $a_i \geq 0$, the amount of time to spend hanging on. The utilities are given by

$$u_i(s_i, s_j, a_i, a_j) = \begin{cases} s_i - a_j & \text{if } a_i > a_j, \\ -a_i & \text{else.} \end{cases} \tag{23}$$

As usual, we replace $[0, \infty)$ with $T = \{\frac{k}{N} : k = 0, 1, \ldots, N^2$, set $t_k = \frac{k}{N}$ and $dt = \frac{1}{N}$.

A. There are always the two asymmetric equilibria, $a_i = t_{N^2}$ and $a_j = 0$, fighting against someone infinitely stubborn, the best response is to quit now.

B. Now suppose that $P(s_i = s_j = v) = 1$ for some $v > 0$, i.e. $F(x) = 1_{[v,\infty)}(x)$.
   (1) Being that the game is symmetric, there is a symmetric equilibrium, and it cannot be a pure strategy equilibrium because the best response to the other dropping out is to stay.
   (2) For a symmetric mixed equilibrium, i cannot put non-infinitesimal mass on quitting at any t_k: this makes j's payoff jump a non-infinitesimal amount by waiting until t_k plus some infinitesimal to quit; so j is not indifferent between t_k and other points, violating the symmetry of the equilibrium.
   (3) At any $h_{|t_k-}$ where neither has quit, the players are in a 2×2 game with payoffs

	Quit	Stay
Quit	$(-t_k, -t_k)$	$(-t_k, v - t_k)$
Stay	$(v - t_k, -t_k)$	(c_k, c_k)

   where c_k is the continuation payoff.
   Letting $\gamma_k \in (0, 1)$ denote the probability of quitting at t_k, we use indifference between Quit and Stay right now for one condition, and the observation that c_k can be had by the indifference condition between Quit and Stay at t_{k+1} to give us

$$\gamma_k(v - t_k) + (1 - \gamma_k)c_k = -t_k, \text{ and } c_k = -t_{k+1} = -(t_k + dt). \tag{24}$$

   Solving yields $\gamma_k = \frac{1}{v+dt}dt$.

Comments: γ_k is independent of k, and we should expect this because the only difference between the game at t_k and t_{k+1} is that we subtract dt from all of the payoffs in the 2×2 matrix; the waiting time until i quits is negative exponential with mean v, which means that the waiting time until the first player quits is negative exponential with mean $v/2$ so that equilibrium payoffs or $(0, 0)$, and we should expect this since the players must be indifferent between any of the times t_k, and 0 is one of the times.

C. Now suppose that F is continuous and has a density f. Verify that the following strategy vector is a symmetric Nash equilibrium: compete until $b(s_i)$ where

$$b(s_i) = \int_0^{s_i} \frac{tf(t)}{1-F(t)}\,dt. \tag{25}$$

D. Let G be the cdf of the quitting times in the previous equilibrium, that is,

$$G(a) = P(b(s_i) \le a) = P\left(\int_0^{s_i} \frac{tf(t)}{1-F(t)}\,dt \le a\right), \tag{26}$$

and suppose that for some $\epsilon > 0$, $F(v-\epsilon) = 0$ and $F(v+\epsilon) = 1$ where $0 < v - \epsilon$. The following steps lead to the conclusion that $G(a) \ge 1 - e^{-a/(v+\epsilon)}$.

a. Show that $G(a) \ge P\left((v+\epsilon)\int_0^{s_i} \frac{f(t)}{1-F(t)}\,dt \le a\right)$.

b. Show that the random variable $F(s_i)$ has the uniform distribution.

c. Show that $\int_0^s \frac{f(t)}{1-F(t)}\,dt = -\log(1-F(s))$.

d. Show that $P\left((v+\epsilon)\int \frac{f(t)}{1-F(t)}\,dt \le a\right) = P(F(s_i) \le 1 - e^{-a/(v+\epsilon)})$. Combining, conclude that

$$G(a) \ge 1 - e^{-a/(v+\epsilon)}. \tag{27}$$

E. Show that

$$G(a) \le 1 - e^{-a/(v-\epsilon)}. \tag{28}$$

F. Let F_n be a sequence of continuous cdfs with density f_n and suppose that for all $\epsilon > 0$, $F_n(v-\epsilon) \to 0$ and $F_n(v+\epsilon) \to 1$. Show that the corresponding cdfs of quitting times, G_n, converge[3] to the symmetric equilibrium quitting time H, for the case that $F(x) = 1_{[v,\infty)}(x)$.

A more general analysis of a war of attrition can be had as follows. Two competitors face each other in a battle of wills, s/he who hangs on the longest wins. We suppose that the value to competitor types, t_i, are iid $U[0,1]$, and the value to type t_i is given by $v(t_i) = F^{-1}(t_i) \ge 0$ where F is a continuous cdf with density f. Denote pure strategies for i as $\sigma_i : [0,1] \to [0,\infty]$ where $\sigma_i(t_i)$ is the time at which type t_i stops fighting. The action $a_i = \infty$ is interpreted as "never stop fighting." The utilities are given by

$$u_i(t_i, t_j, \sigma_i, \sigma_j) = \begin{cases} v(t_i) - \sigma_j(t_j) & \text{if } \sigma_i(t_i) > \sigma_j(t_j), \\ -\sigma_i(t_i) & \text{else.} \end{cases} \tag{29}$$

1. Show that if i uses a non-decreasing strategy, then j has a response in the non-decreasing strategies.
2. Show that the non-decreasing functions from $[0,1]$ to $[0,\infty]$ are a complete lattice.
3. Show that the set of equilibria in non-decreasing strategies has a lattice structure.

[3] In the usual sense of convergence of cdf's, $G_n(a) \to H(a)$ for all continuity points of $H(\cdot)$.

It is easy to verify that $\sigma_i(t_i) \equiv 0$ and $\sigma_j(t_j) \equiv \infty$ is an asymmetric equilibrium. The rest of the problem focuses on the properties of symmetric equilibria in non-decreasing strategies, $(\sigma_i, \sigma_j) = (\sigma, \sigma)$. For non-decreasing strategies, $\sigma_i^{-1} : [0, \infty] \to [0, 1]$ gives the cdf of i's quitting time, and if σ^{-1} is differentiable, then at time a it has hazard rate $h_\sigma(a) = \frac{d\sigma^{-1}(a)/da}{1-\sigma^{-1}(a)}$.

1. If t_i competes until time $a \in [0, \infty]$, show that their payoff is

$$U_i(t_i, a; \sigma) = \int_0^a (v(t_i) - s)\, d\sigma^{-1}(s) - a(1 - \sigma^{-1}(a)). \tag{30}$$

   From this, derive the first order conditions for $a^*(t_i)$ being $\frac{1}{v(t_i)} = h_\sigma(a)$.
2. Show that the equilibrium hazard rate of the duration of the conflict must be non-decreasing. [Note that $a^*(t_i) = \sigma(t_i)$ so that $v(t_i) = v(\sigma^{-1}(a))$.]
3. Argue that $\sigma(0) = 0$ must hold in equilibrium and deduce that $\sigma(t_i) = \int_0^{t_i} \frac{v(s)}{1-s}\, ds$. From this, conclude that the optimal strategy increases in $v(\cdot)$. Examine the possibility of deriving this from supermodularity in eqn. (30).

4.4. **Brownian Monitoring.** standard parts of graphs.

4.5. **Poisson Monitoring.** standard parts of graphs.

4.6. **Continuous Time Martingales.** Urn models.

4.7. **Itô's Lemma.** Trading models: Kreps; Harrison-Kreps; Harrison-Pliska.

5. Standard and Nonstandard Superstructures

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, or near enough to internal that we can assign probabilities to them. At this point, it makes sense to go back and be a bit more clear about what we meant by "putting *'s on everything."

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 et al., 2009] 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 5.1.1. *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 5.1.2. *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 et al., 2009]

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 5.3.1. *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 $G = \langle G, G, G, \dots \rangle$, then G is **standard**, otherwise it is **internal**.
   a. $\langle [0,1], [0,1], [0,1], \dots \rangle$ is the standard set we denote ${}^*[0,1]$, ${}^*\mathbb{R}_+ = \langle \mathbb{R}_+, \mathbb{R}_+, \mathbb{R}_+, \dots \rangle$. ${}^*[0,1]$ contains the standard point $\langle r, r, r, \dots \rangle$ as long as $0 \le r \le 1$, ${}^*\mathbb{R}_+$ contain unlimited points such as the factorials $\langle n! \rangle$. ${}^*[0,1]$ also contains the infinitesimal $\langle 1, \frac{1}{2}, \frac{1}{3}, \dots \rangle$, and the nearstandard point $\langle r+1, r+\frac{1}{2}, r+\frac{1}{3}, \dots \rangle$ as long as $0 \le r < 1$.
   b. $F = \langle \{0,1\}, \{0, \frac{1}{2}, 1\}, \{0, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, 1\}, \dots$ is an internal set satisfying $d_H(F, {}^*[0,1]) = \langle \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots \rangle \simeq 0$. 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 $K_a = \langle K_{a,1}, K_{a,2}, \dots \rangle$ and $K_b = \langle K_{b,1}, K_{b,2}, \dots \rangle$ in ${}^*\mathcal{K}_{\mathbb{R}}$, we have set things up so that $d_H(K_a, K_b) = \langle d_H(K_{a,1}, K_{b,1}), d_B(K_{a,2}, K_{b,2}), \dots \rangle$.
   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 $(\Omega, \mathcal{F}, P) = \langle (\Omega_1, \mathcal{F}_1, P_1), (\Omega_2, \mathcal{F}_2, P_2), (\Omega_3, \mathcal{F}_3, P_3) \dots \rangle$, then $\mathbb{R}^\Omega = \langle \mathbb{R}_1^\Omega, \mathbb{R}_2^\Omega, \mathbb{R}_3^\Omega, \dots \rangle$ is the set of *-random variables on the internal set $\Omega = \langle \Omega_1, \Omega_2, \Omega_3, \dots \rangle$.

Some more examples.

Example 5.4.1. *${}^*\mathbb{N}$ is standard while $F = \langle \{k/2^n : k = 0, \dots, n \cdot 2^n\} \rangle$ 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 $Poly = \langle \text{span}(\{x^k : k = 0, \dots n\}) \rangle$ is an internal subset of ${}^*C([0,1])$, and the Stone-Weierstrass theorem tells us that every standard $f \in {}^*C([0,1])$, $d(f, Poly) \simeq 0$.*

5.5. **Internal Sets for Stochastic Processes.** To recognize when we have an internal set, it is useful to know when we don't.

5.6. **Some External Sets.**

Theorem 5.6.1. *The following sets are external.*

a. $A_1 = \{n \in {}^*\mathbb{N} : n \text{ is standard}\}$.
b. $A_2 = \{n \in {}^*\mathbb{N} : n \text{ is nonstandard}\}$.
c. $A_3 = \{r \in {}^*\mathbb{R} : r \text{ is limited}\}$.
d. $A_4 = \{r \in {}^*\mathbb{R} : r \text{ is unlimited}\}$.
e. $A_5 = \{r \in {}^*\mathbb{R} : r \text{ is infinitesimal}\}$.

Proof. If $A_1 = \langle A_{1,1}, A_{1,2}, A_{1,3}, \ldots\rangle$, set $a_n = \max A_{1,n}$ and let $a = \langle a_1, a_2, \ldots\rangle$. For any limited $n \in \mathbb{N}$, $a > n$, and $a \in A_1$. If A_2 is internal, then so is A_2^c, but $A_2^c = A_1$.

If $A_4 = \langle |A_{4,1}|, |A_{4,2}|, |A_{4,3}|, \ldots\rangle$, set $a_n = \inf |A_{4,n}|$ and let $a = \langle a_1, a_2, \ldots\rangle$. It cannot be that a is limited because this would mean that $|A_4| \cap (a, a+1) \neq \emptyset$ so that $|A_4|$ contains a limited number. But if a is unlimited, then so is $a/2$, and $a/2 < a$ implying that A_4 is missing some unlimited numbers. A_3 is the complement of A_4 so cannot be internal either. If A_5 is internal, then so is $\{1/|x| : x \in A_5, x \neq 0\}$, but this is the set of all unlimited positive numbers, which cannot be internal by the same arguments. □

Corollary 5.6.1.1. *If an internal subset of* ${}^*\mathbb{R}_+$ *or* ${}^*\mathbb{N}$ *contains arbitrarily small unlimited numbers, then it contains a limited number. If an internal subset of* ${}^*\mathbb{R}_+$ *contains arbitrarily large infinitesimals, then it contains a limited non-zero number.*

Here is an implication that will be useful many times.

Lemma 5.6.2 (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 $n \leq m$.*

Proof. Consider the set $S := \{m \in {}^*\mathbb{N} : (\forall n \leq m)[|x_n| < 1/m]\}$. This is an internal set (which you should figure out how to check), and contains arbitrarily large integers, hence contains an infinite integer, m. □

Lemma 5.6.3. *If n is limited and for each $i \leq n$, $x_i \simeq y_i$, then $\sum_{i=1}^n x_i \simeq \sum_{i=1}^n y_i$.*

Proof. $|\sum_{i=1}^n x_i - \sum_{i=1}^n y_i| \leq \sum_{i=1}^n |x_i - y_i| \leq n \max\{|x_i - y_i| : i \leq n\} \simeq 0$. □

5.7. **Statements and the Transfer Principle.** 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 $V({}^*S)$.

Example 5.7.1. *Let us examine the statement that subsets of $\mathbb{R}$ that are bounded above have a supremum. The following statement is true in $V(S)$:*

$$(\forall A \in \mathcal{P}(\mathbb{R})[\mathbb{B}(A) \Rightarrow \mathbb{S}(A)] \tag{31}$$

where $\mathbb{B}(A)$ is the statement "A is bounded above" and $\mathbb{S}(A)$ is the statement "A has a supremum." If $B \subset \mathcal{P}(\mathbb{R})$ is the class of bounded sets and $S \subset \mathcal{P}$ is the class of sets having a supremum, this is $B \subset S$. The statement in $V({}^*S)$ is

$$(\forall A \in {}^*\mathcal{P}(\mathbb{R})[{}^*\mathbb{B}(A) \Rightarrow {}^*\mathbb{S}(A)], \tag{32}$$

or ${}^*B \subset {}^*S$.

In more detail, $\mathbb{B}(A)$ is the statement that

$$(\exists B \in \mathbb{R})(\forall a \in A)[a \leq B]. \tag{33}$$

The *'d version is

$$(\exists B \in {}^*\mathbb{R})(\forall a \in A)[a \leq B] \tag{34}$$

where the "$\in$" and the "$\leq$" maybe ought to have *'s as well.

*Note that the part "$(\forall A \in {}^*\mathcal{P}(\mathbb{R})$" means that A needs to be an internal set. This yields another proof that the set of infinitesimals is **not** internal: the class of infinitesimal is certainly bounded above, e.g. by 1; if $s \in {}^*\mathbb{R}$ is its supremum, it is either infinitesimal, in which case $2s$ is also infinitesimal and s was not the supremum, or s is not infinitesimal, in which case $s/2$ is not infinitesimal but is an upper bound for the set of infinitesimals.*

6. Some Real Analysis

Continuity, compactness, uniform continuity, the standard part map, the Theorem of the Maximum, limit games and limit equilibria, the Riesz representation theorem, Glicksberg-Fan fixed point theorems, more equilibrium refinement for compact and continuous games, infinite signaling games and cheap talk theorems, infinite extensive form games.

6.1. **Closed Sets and Closure.** Internal sets have a form of the finite intersection property. Implications of this are many.

6.1.1. *The standard part mapping.*

6.1.2. *Closedness of Refined Sets of Equilibria.*

6.2. **Continuity and Uniform Continuity.**

6.2.1. *$C(X;\mathbb{R})$, X compact.* $f(x) = x^2$, $x \in M$ versus $x \in {}^*M$.

6.2.2. *Near continuity.*

6.2.3. *The Riemann-Stieltjes Integral.*

6.2.4. *Some near interval control theory.* Include sufficient conditions for near interval solution to have a continuous standard part [Fleming and Rishel, 1975, Ch. 1] plus stuff from [Clarke, 2013]. Basically, concavity of $U(t, x, \dot{x})$ in its third argument.

6.3. **Theorem of the Maximum.**

6.3.1. *In Control Theory.* Re-prove closedness of eq'm sets, continuity of $t \mapsto U(t, x^*(t), \dot{x}^*(t))$ in control theory.

6.3.2. *Single person problems.*

6.3.3. *Limit games and limit equilibria.*

6.4. **Compactness.** Robinson's theorem.

6.4.1. *Existence of optima.*

6.4.2. *Existence of equilibria.*

6.4.3. *Existence of extened equilibrium outcomes.* Material from finitistic games work goes here.

6.4.4. *Compact sets of probabilities on $\mathbb{R}$.* $d_H(F_n, F) \to 0$ iff weak convergence. The robustnik interpretation.

Ulam's theorem.

Tightness = compactness.

The Gaussian CLT.

The problem of moments and distributions infinitesimally close to the standard normal.

6.5. **Probabilities on Metric Spaces.**

6.5.1. *Loeb Measures.*

6.5.2. *Riesz representation theorem.*

6.5.3. *Denseness of finitely supported probabilities.*

6.5.4. *Tightness and compactness.*

6.6. **Derivatives.**

6.6.1. *Basics.*

A. 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 $\lim_{x \to 0} \frac{f(x)}{g(x)} \simeq \frac{f(dx)}{g(dx)} \simeq \frac{f'(0)}{g'(0)}$.

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

C. 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 $\{{}^\circ r : r \in {}^*\mathbb{R}, r \text{ finite}\} = \mathbb{R}$.

6.6.2. *The implicit function theorem.* Uses, proof.

6.6.3. *Lebesgue's density theorem.* Use $\text{st}^{-1}(A)$ characterization of measurable sets ...

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

A. $n \mapsto s_n$ is a Cauchy sequence in $\mathbb{R}$ iff ${}^*s_n \simeq {}^*s_m$ for all $n, m \in {}^*\mathbb{N} \setminus \mathbb{N}$.

B. 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 $c \in (0,1)$.

6.8. **A Duality Approach to Patience.** We begin with the overview that will only make sense if you already know a fair amount of functional analysis. If you are not such a person, it would be a good come back to this after each of the subsequent developments.

> Let $(\mathfrak{X}, \|\cdot\|)$ be a normed vector algebra of $\mathbb{R}$-valued functions containing the constant function $\mathbf{1}$, and partially ordered by $\leq$; let $(\mathfrak{X}^\dagger, \|\cdot\|^\dagger)$ be the dual space of $\mathfrak{X}$ with the associated dual norm and partial order $x^\dagger \geq 0$ if $\langle x, x^\dagger \rangle \geq 0$ for all $x \geq 0$; the "duality approach" begins with a linear subspace, $\mathfrak{L} \subset \mathfrak{X}$, uses this to define a subset $\mathfrak{D}_\mathfrak{L}$ of non-negative elements of $\mathfrak{X}^\dagger$ having norm 1 and annihilating $\mathfrak{L}$, specifically, $\mathfrak{D}_\mathfrak{L} = \{x^\dagger \in \mathfrak{X}^\dagger : x^\dagger \geq 0, x^\dagger(\mathbf{1}) = 1, x^\dagger(\mathfrak{L}) = 0\}$; with this set, we define the concave, homogenous of degree 1 utility function $\mathcal{U}_\mathfrak{L}(x) = \min\{\langle x, x^\dagger \rangle : x^\dagger \in \mathfrak{D}_\mathfrak{L}\}$; this automatically has the property that $[(x - y) \in \mathfrak{L}] \Rightarrow [\mathcal{U}_\mathfrak{L}(x) = \mathcal{U}_L(y)]$; alternatively, one can start with a set $\mathfrak{D}$, of non-negative, norm 1 elements of $\mathfrak{X}^\dagger$, define $\mathfrak{L}_\mathfrak{D} = \mathfrak{D}^\perp$, and observe that $\mathcal{U}_{\mathfrak{L}_\mathfrak{D}}(x) = \min\{\langle x, x^\dagger \rangle : x^\dagger \in \mathfrak{D}\}$; when $\mathfrak{X}$ is the set of bounded sequences of utilities, different choices of $\mathfrak{L}$ having to do with long run behavior give different definitions of patience.

6.8.1. *Preliminaries.* Bounded sequences of utilities, $(u_t)_{t=1}^\infty$ belong to $\ell_\infty := \{u \in \mathbb{R}^\mathbb{N} : (\exists B \in \mathbb{R}_+)(\forall n \in \mathbb{N})[\,|u_n| \leq B\,]\}$, and we use the sup-norm, $\|u\| = \sup_t |u_t|$. This means that $(\ell_\infty, \|\cdot\|)$, is the Banach space $(C_b(\mathbb{N}), \|\cdot\|)$. We are interested in the properties of "patient" and "time invariant" preferences over ℓ_∞. We first gather the pertinent definitions.

Definition 6.8.1. *A (complete transitive) preference relation, $\succsim$, on ℓ_∞ is **continuous** if it can be represented by a continuous utility function, it is monotonic if $[u \geq v] \Rightarrow [u \succsim v]$, and it respects intertemporal smoothing if for all $u \sim v \in \ell_\infty$ and all $\alpha \in (0,1)$, $\alpha u + (1-\alpha)v \succsim u$.*

The last condition gives quasi-concavity of the utility function, but the kinds of preferences we will be looking at can all be represented by a *concave* function on ℓ_∞ that is homogenous of degree 1.

6.8.2. *Infinite Patience.* There are various ways to approach the idea of a preference ordering on ℓ_∞ being infinitely patient that can be expressed using some useful linear subspaces, $\mathcal{L}$, of ℓ_∞. The idea will be that $[(u - v) \in \mathcal{L}] \Rightarrow [u \sim v]$, which means that the larger is $\mathcal{L}$, the more restrictive the condition. In the following, F is mnemonic for Finite, K is mnemonic for Kronecker, and A is mnemonic for Average:

$\mathcal{L}_F = \{u \in \ell_\infty : u_t \neq 0$ for only finitely many $t\}$;
$\mathcal{L}_K = \{u \in \ell_\infty : \sum_{t=1}^\infty \frac{u_t}{t}$ exists$\}$; and
$\mathcal{L}_A = \{u \in \ell_\infty : \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^T u_t = 0\,\}$.

We will show that $\mathcal{L}_F \subsetneq \mathcal{L}_K \subsetneq \mathcal{L}_A \subsetneq \mathcal{L}_T$ where $\mathcal{L}_T$ is as of yet, undefined.

Definition 6.8.2. *We have the following kinds of patience:*

(a) *A preference relation is F-**patient** if* $[(u-v) \in \mathcal{L}_F] \Rightarrow [u \sim v]$*, that is, if any changes on any finite set of time indexes leaves the decision maker indifferent.*
(b) *A more restrictive condition is K-**patience**, which is* $[(u-v) \in \mathcal{L}_K] \Rightarrow [u \sim v]$*, so every F-patient preference relation is K-patient because* $\mathcal{L}_F \subset \mathcal{L}_K$*.*
(c) *A yet more restrictive condition is A-**patience**, which is* $[(u-v) \in \mathcal{L}_A] \Rightarrow [u \sim v]$*, so every K-patient preference relation is A-patient because* $\mathcal{L}_K \subset \mathcal{L}_A$*.*
(d) *The most restrictive condition is T-**patience**, which replaces* $\mathcal{L}_A$ *with the as-yet-undefined larger linear subspace* $\mathcal{L}_T$*.*

Let $(\mathfrak{X}, \|\cdot\|)$ be a normed linear space like $(\ell_\infty, \|\cdot\|_\infty)$ and let $(\mathfrak{X}^\dagger, \|\cdot\|)$ be the space of continuous linear functionals on $\mathfrak{X}$ with the norm $\|x^\dagger\| = \sup\{|x^\dagger(x)| : \|x\| \leq 1\}$. Recall that the kernel of a linear mapping, $x^\dagger$, is the set of x such that $x^\dagger(x) = 0$. Further, given a subset of $\mathfrak{L} \subset X$, its "perpendicular complement" is $\mathfrak{L}^\perp = \{x^\dagger \in \mathfrak{X}^\dagger : x^\dagger(\mathfrak{L}) = 0\}$, that is, the set of continuous linear functionals that have $\mathfrak{L}$ as a subset of their kernel.

We can now say what the larger linear subspace is: $\mathcal{L}_T$ consists of the u's in the kernel of every Translation invarient probability (and all we will need to do is to define translation invariant probabilities). The harder part of the following is the second inclusion, which is a version of **Kronecker's lemma**, the example showing the inclusion is strict is rather painful.

Lemma 6.8.3. *$\mathcal{L}_F \subsetneq \mathcal{L}_K \subsetneq \mathcal{L}_A$, all three are all linear subspaces, and $\mathcal{L}_A^\perp \subsetneq \mathcal{L}_K^\perp \subsetneq \mathcal{L}_F^\perp$.*

Proof. The second part, "all three are all linear subspaces," is immediate.
$\mathcal{L}_F \subsetneq \mathcal{L}_K$: If $u \in \mathcal{L}_F$, then $u_t = 0$ for all sufficiently large t, hence $u \in \mathcal{L}_K$. Taking e.g. $u_t = 1/t$ gives an element of $\mathcal{L}_K$ that does not belong to $\mathcal{L}_F$.
$\mathcal{L}_K \subsetneq \mathcal{L}_:$: Suppose now that $u \in \mathcal{L}_K$ and pick an arbitrary infinite T'. We must show that $\frac{1}{T'}\sum_{t=1}^{T'} {}^*u_t \simeq 0$. Because $u \in \mathcal{K}$, we know that for all infinite $T' > T$, $\sum_{t=1}^{T'} \frac{{}^*u_t}{t} \simeq \sum_{t=1}^{T} \frac{{}^*u_t}{t}$ and both are nearstandard. Being infinitely close to each other, $\sum_{t=T+1}^{T'} \frac{{}^*u_t}{t} \simeq 0$. Consider the internal set $\mathbb{T} := \{\tau : \frac{1}{T'}\sum_{t=1}^{\tau} |{}^*u_t| < \frac{1}{\tau}$. Because T' is infinite and $u \in \ell_\infty$, $\mathbb{T}$ contains arbitrarily large finite elements, hence by overspill, an infinite element, call it T.

$$\frac{1}{T'}\sum_{t=1}^{T'} {}^*u_t = \frac{1}{T'}\left[\sum_{t=1}^{T} {}^*u_t\right] + \frac{1}{T'}\left[\sum_{t=T+1}^{T'} {}^*u_t\right]. \tag{35}$$

Now, the first term is infinitesimal because its absolute value is bounded above by $\frac{1}{T'}\sum_{t=1}^{T} |{}^*u_t| < \frac{1}{T} \simeq 0$. For the second term, we know that $\sum_{t=T+1}^{T'} \frac{{}^*u_t}{t} \simeq 0$ and for each $t \in \{T+1, \ldots, T'\}$, $\frac{1}{T'} \leq \frac{1}{t}$.

To show the inclusion is strict, we will check that (i) $u_t = \min\{1, 1/\log(t)\}$ does not belong to $\mathcal{L}_K$, i.e. $\sum_{t=2}^{T} \frac{1}{t\log(t)} \simeq \infty$ for any infinite T, but (ii) $\frac{1}{T}\sum_{t=2}^{T} \frac{1}{\log(t)} \simeq 0$ for any infinite T.
(i) Consider the integral $\int_2^T \frac{1}{x\log(x)}\,dx$. Using the change of variable $y = \log(x)$, we have $dy = \frac{1}{x}dx$ so that $\int_2^T \frac{1}{x\log(x)}\,dx = \int_{\log(2)}^{\log(T)} \frac{1}{y}\,dy$, and this is $\log(\log(T))$ minus a constant, hence goes to 0 (albeit very slowly).

(ii) Consider $\frac{1}{T}$ times the integral $\int_2^T \frac{1}{\log(x)}\,dx$. The integral has a nasty form, $\log(\log(T))+\log(T)+\sum_{k=2}^{\infty}\frac{\log(T)^k}{k\,k!}$ minus a constant. Now $\frac{\log(T)}{T}\to 0$ so that $\frac{\log(\log(T))}{T}\to 0$, so all that we need to show is that $\frac{1}{T}$ times the last term goes to 0. Ignoring the constant, we can use l'Hôpital's rule, which tells us that $\lim_T \frac{\sum_{k=2}^{\infty}\frac{\log(T)^k}{k\,k!}}{T}$ is the limit of the derivatives top and bottom,

$$\lim_T \frac{\sum_{k=2}^{\infty}\frac{\log(T)^{k-1}(1/T)}{k!}}{1} = \lim_T \frac{e^{\log(T)}-(1+\log(T))}{T\log(T)} = \lim_T \frac{T-1-\log(T)}{T\log(T)} = 0. \quad (36)$$

Finally, to show that $\mathcal{L}_A^{\perp}\subsetneq\mathcal{L}_K^{\perp}\subsetneq\mathcal{L}_F^{\perp}$, it is sufficient to show that the strict inclusion relations holds for the closurures, $\overline{\mathcal{L}}_F\subsetneq\overline{\mathcal{L}}_K\subsetneq\overline{\mathcal{L}}_A$.

To see that there is an element of $\mathcal{L}_K$ at distance 1 from every element of $\overline{\mathcal{L}}_F$, let A be the infinite set $\{t^2 : t\in\mathbb{N}\}$, set $u=1_A$, and note that $\sum_t \frac{u_t}{t}=\sum_t\frac{1}{t^2}<\infty$.
To see that there is an element of $\mathcal{L}_F$ at distance 1 from every element of $\overline{\mathcal{L}}_A$, let $B=\{\lfloor t\cdot\log(t)\rfloor : t\in\mathbb{N}, t\geq e^2\}$ and set $u=1_B$. For large enough τ, $\sum_{t\geq\tau}\frac{u_t}{t}\geq 0.9\sum_{t=\tau}^{\infty}\frac{1}{t\log(t)}=\infty$ while $\sum_{t=2}^{T}u_t\leq\sum_{t=2}^{T}\frac{1}{\log(t)}$. □

6.8.3. *Continuous Linear Functionals.* We will study monotonic preferences that that can be represented by continuous concave functions on ℓ_∞. Continuous concave functions are the lower envelope of the continuous affine functions that majorize them. This makes knowing how to represent continuous linear functions a crucial first step.

The dual space of ℓ_∞ is denoted $\ell_\infty^\dagger$. It is strictly larger than $\ell_1 := \{p\in\mathbb{R}^{\mathbb{N}} : \sum_t |p_t|<\infty\}$, where each $p\in\ell_1$ defines/is identified with an element of $\ell_\infty^\dagger$ by $f_p(u)=\langle u,p\rangle=\sum_t u_t p_t$.

Example 6.8.1. *For $\delta\simeq 1$ pick $T\in{}^*\mathbb{N}\backslash\mathbb{N}$ such that $(1-\delta)\sum_{t=1}^{T}\delta^{t-1}\simeq 1$ and consider the preference relation represented by the function $f(u)={}^\circ(1-\delta)\sum_{t=1}^{T}u_t\delta^{t-1}\simeq 1$. For $T\in{}^*\mathbb{N}\setminus\mathbb{N}$, consider the preference relation represented by the function $g(u)={}^\circ\frac{1}{T}\sum_{t=1}^{T}u_t$. These are continuous linear functionals on ℓ_∞, i.e. $f,g\in\ell_\infty^\dagger$, and neither is an element of ℓ_1 as can be seen by considering the unit vectors $e_k=(u_t)_{t=1}^{\infty}$ with $u_k=1$, and $u_t=0$ for $t\neq k$. For each $k\in\mathbb{N}$, $f(e_k)=g(e_k)=0$, but $\langle e_k,p\rangle=p_k$, and this is equal to 0 for all $k\in\mathbb{N}$ iff $p=0$ in ℓ_1.*

All continuous linear functionals on ℓ_∞ can be represented as the standard part of inner products like those in this example.

Theorem 6.8.4. *If $f\in\ell_\infty^\dagger$, then there exists a star-finite set $\{1,\dots,T\}\subset{}^*\mathbb{N}$ and $\eta_1,\dots,\eta_T$ with $\sum_{t=1}^{T}|\eta_t|$ limited and $f(u)={}^\circ\langle{}^*u,\eta\rangle$.*

Proof. Let $\mathcal{N}$ denote the class of all subsets of $\mathbb{N}$. For any $u\in\ell_\infty$ and any $n\in\mathbb{N}$, consider the two sequences of simple functions,

$$\overline{U}^n(t) := \sum_{k=-\infty}^{+\infty}\frac{k}{2^n}1_{u^{-1}((\frac{k-1}{2^n},\frac{k}{2^n}])}(t), \text{ and } \underline{U}^n(t) := \sum_{k=-\infty}^{+\infty}\frac{k-1}{2^n}1_{u^{-1}((\frac{k-1}{2^n},\frac{k}{2^n}])}(t). \quad (37)$$

We have $\underline{U}^n<u\leq\overline{U}^n$ and $\|\overline{U}^n-\underline{U}\|\leq\frac{1}{2^n}$ so that either of these classes of simple functions is dense in ℓ_∞.

If $f \in \ell_\infty^\dagger$, then it is Lipschitz with Lipschitz constant $\|f\|^\dagger$, hence determined by its values on any dense subset, hence determined by its values on the class of simple functions. For any finite set of simple functions, $\mathcal{U} := \{U_a : a = 1, \ldots, A\}$, let $\mathcal{P}_\mathcal{U}$ denote the partition of $\mathbb{N}$ generated by sets of the form $U_a^{-1}(r)$, $r \in \mathbb{R}$. By basic linear algebra, there exists a finitely supported measure, $\eta = \eta(\mathcal{U})$, on $\mathbb{N}$ with $f(U_a) = \sum_{t \in \mathbb{N}} U_a(t)\eta(t)$ and $\sum_t |\eta(t)| \leq \|f\|^\dagger$. Let $\mathcal{U}'$ be an exhaustive, *-finite set of *-simple functions, and let $\eta = \eta(\mathcal{U}')$. □

Another way to understand this theorem is to recall that $\ell_\infty = C_b(\mathbb{N})$, the Riesz representation theorem tells us that the dual space of $C_b(\mathbb{N})$ is the set of bounded, finitely additive measures on $\mathbb{N}$, and this result is telling us that such measures have a star-finite representation.

6.8.4. *Star-finitely Supported Probabilities.* For our purposes, the probability measures are enough.

Definition 6.8.5. *$\eta \in {}^*\ell_1$ is a **star-finite probability (sfp)** if $\eta \geq 0$, $\sum_t \eta_t \simeq 1$, and $\{t \in {}^*\mathbb{N} : \eta_t > 0\}$ is star-finite. An sfp η is **remote** $\eta(A) \simeq 0$ for all finite A.*

Let $L(\eta)$ denote the Loeb measure generated by η. Given that $L(\eta)$ is countably additive, $\eta(A) \simeq 0$ for all finite A is equivalent to $L(\eta)(\mathbb{N}) = 0$. It turns out that the remote sfp's are the ones that let us get at F-patience.

Lemma 6.8.6. *An sfp η is remote iff for all $u, v \in \ell_\infty$ that differ in only finitely many time periods, $\langle {}^*u, \eta\rangle \simeq \langle {}^*v, \eta\rangle$.*

Remember, u and v differing in only finitely many time periods is $(u - v) \in \mathcal{L}_F$.

Proof. For any finite $A \subset \mathbb{N}$, let $u = 1_A$ and set $v = 0$ so that ${}^\circ\langle {}^*u, \eta\rangle \simeq \langle {}^*v, \eta\rangle = L(\eta)(A) = 0$. Since $\mathbb{N}$ is the countable union of finite sets and $L(\eta)$ is countably additive, $L(\eta)(\mathbb{N}) = 0$. □

6.8.5. *Preferences with a* lim inf *Representation.* Every concave function is the lower envelope of the affine functions that majorize it, to put it another way, if $\mathcal{A}$ is a set of affine functions and $g(u) = \inf_{a \in \mathcal{A}} a(u)$, then $g(\cdot)$ is concave. If we majorize using linear functionals, we get functions that are concave and homogenous of degree 1 (hd1). The following is the first of our interesting **concave** hd1 functionals on ℓ_∞.

Theorem 6.8.7. *For $u \in \ell_\infty$, $\liminf_{t\to\infty} u_t = \inf_{\eta \in R} {}^\circ\langle {}^*u, \eta\rangle$ where R is the set of remote sfp's.*

We will see in just a little bit that we can replace "inf" with "min" in this result.

Proof. Let $\underline{u} = \liminf_{t\to\infty} u_t$. For any remote sfp, ${}^\circ\langle {}^*u, \eta\rangle \geq \underline{u}$, so it is sufficient to find a remote sfp with ${}^\circ\langle {}^*u, \eta\rangle = \underline{u}$.

For each $n \in \mathbb{N}$, let $\mathbb{T}_n$ be the internal set $\{t \in {}^*\mathbb{N} : t \geq n, {}^*u_t < \underline{u} + \frac{1}{n}\}$ so that $\mathbb{T}_n \supset \mathbb{T}_{n-1}$ and $\mathbb{T}_n \neq \emptyset$. By the internal extension principle, the mapping $n \mapsto \mathbb{T}_n$ has an internal extension to a mapping from ${}^*\mathbb{N}$ to ${}^*\mathcal{P}(\mathbb{N})$. For that extension, the internal set N of all n such that $\mathbb{T}_n \supset \mathbb{T}_{n-1}$ and $\mathbb{T}_n \neq \emptyset$ contains all finite n, hence contains an infinite n'. Because any $t \in \mathbb{T}_{n'}$ is greater than or equal to n', any $\eta \in \Delta(\mathbb{T}_{n'})$ is remote, and for any such η, $|\langle {}^*u, \eta\rangle - \underline{u}| \leq \frac{1}{n'} \simeq 0$. □

An Interpretation: If one values a sequence u by $\langle {}^*u, \eta\rangle$ for a remote sfp, then one cares only about the far future. Using the utility function $\underline{l}(u) = \liminf_{t\to\infty} u_t$ to judge rewards corresponds to judging u by its worst behavior in the far future. This seems unreasonably pessimistic.

Example 6.8.2. *Let u be the sequence that begins with a single $+1$, then has its next 2^2 entries being -1, then has a single $+1$, then has the next 3^3 entries being -1, then has a single $+1$, then has the next 4^4 entries being -1. This u is overwhelmingly often negative while $v := -u$ is overwhelmingly often positive, yet $\underline{l}(u) = \underline{l}(v) = -1$.*

One way to get around such examples is the pay attention to the long run average of the payoffs, and we will turn to this very soon.

Another Interpretation: Let Δ^{fa} denote the set of finitely additive probabilities on the integers. A sub-basis for the weak* topology on Δ^{fa} is given by sets of the form $G(\mu : u, \epsilon) := \{\nu \in \Delta^{fa} : |\int_{\mathbb{N}} u\, d\nu - \int_{\mathbb{N}} u\, d\mu| < \epsilon\}$, $u \in C_b(\mathbb{N})$, $\epsilon > 0$. This means that all weak*-open sets are unions of finite intersections of such sets, and $\mu_\alpha \to \mu$ iff for all $u \in C_b(\mathbb{N})$, $\int u\, d\mu_\alpha \to \int u\, d\mu$. The next Theorem and Corollary will show that $\underline{l}(u) = \min_{\mu\in\mathcal{R}}\langle u, \mu\rangle$ where $\mathcal{R} = \mathrm{st}\,(R)$. This kind of preference has a Choquet integral representation. For $E \subset \mathbb{N}$, define $c(E) = \min\{\mu(E) : \mu \in R\}$.

Lemma 6.8.8. *For all $u \in \ell_\infty$, $\underline{l}(u) = \int u\, dc$ where the integral is in the sense of Choquet.*

The following is a famous consequence of Alaoglu's theorem.

Theorem 6.8.9. *Δ^{fa} is compact in the weak*-topology.*

Proof. Let $\eta \in {}^*\Delta^{fa}$. We must show that $\mathrm{st}\,(\eta) \in \Delta^{fa}$. But it is immediate that ν defined by $\nu(A) = \mathrm{st}\,(\eta)(A)$ belongs to Δ^{fa}. Finally, by linearity and the $\|\cdot\|$-denseness of the simple functions, $\nu = \mathrm{st}\,(\eta)$. □

Note that $\Delta^{ca}(X)$ is **not** weak* compact unless X is a well-behaverd compact space.

Definition 6.8.10. *$\mu \in \Delta^{fa}$ is **weightless** or **purely finitely additive** if for all finite $A \subset \mathbb{N}$, $\mu(A) = 0$.*

Here is the relation between weightless probabilities and the remote sfp's.

Corollary 6.8.10.1. *$\mathcal{R} = \mathrm{st}\,(R)$ is a compact convex subset of Δ^{fa} and $\mu \in \mathcal{R}$ iff μ is weightless.*

Proof. Finite intersection property. □

6.8.6. *Concave F-Patient Preferences.* Being the infimum of a collection of linear functionals on ℓ_∞, the utility function $\underline{l}(u) = \liminf_{t\to\infty} u_t$ is concave. In this context, concavity indicates a preference for intertemporal smoothing of utilities, i.e. $\underline{l}(\alpha u + (1-\alpha)v)) \geq \alpha\underline{l}(u) + (1-\alpha)\underline{l}(v) \geq \min\{\underline{l}(u), \underline{l}(v)\}$.

We can turn the analysis on its head a bit too: we just showed that having tangents that are remote sfp's means that we respect F-patience; being concave and respecting F-patience means that the tangents are, up to a scalar factor, remote sfp's.

Theorem 6.8.11. *If $\mathcal{V} : \ell_\infty \to \mathbb{R}$ is continuous, monotonic, concave and F-patient, then for any $g \in D\mathcal{V}(u)$, $g = \lambda f$ for some $\lambda \geq 0$ and remote sfp f.*

Proof. Straightforward. □

6.8.7. *Preferences with a* $\liminf$*-Average Representation.*

Definition 6.8.12. *A remote sfp is of* ***Cesaro type*** *if* $\eta_1 = \cdots = \eta_T = \frac{1}{T}$ *for* $T \in {}^*\mathbb{N} \setminus T$.

Theorem 6.8.13. *For* $u \in \ell_\infty$, $\underline{L}(u) := \liminf_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} u_t = \inf_{\eta\in\mathcal{C}} {}^\circ\langle {}^*u, \eta\rangle$ *where* $\mathcal{C}$ *is the set of remote sfp's of Cesaro type.*

Proof. Let T_k be a sequence with $\frac{1}{T_k}\sum_{t=1}^{T_k} u_t \to \underline{L}(u)$. Let T be the equivalence class of the sequence $(T_1, T_2, \ldots)$, and let η_T be the uniform distribution on $\{1, \ldots, T\}$. It is immediate that $\langle {}^*u, \eta_T\rangle \simeq \underline{U}$, and no η of Cesaro type can have a lower integral against ${}^*u$. □

It is possible in ℓ_∞, that on average, one is living in good times, but that there will be infinitely many bad times of arbitrarily long length. One could be more pessimistic than the $\underline{L}(\cdot)$ utility function by overwhelmingly paying attention to those longer and longer stretches of bad news.

Example 6.8.3. *Let* u *have the first* 2^2 *entries being* $+1$*, next* 2 *entries being* -1*, then next* 3^3 *entries being* $+1$*, then next* 3 *entries being* -1*, the next* 4^4 *entries being* $+1$*, the next* 4 *entries being* -1*, etc.*

6.9. **HERE LIE DRAGONS.** It is worth bearing in mind the observation (due to Choquet?) that a Banach space is separable if and only if one can define a strictly concave $\mathbb{R}$-valued function on it. The space ℓ_∞ is **not** separable, and the concave functions that we have been looking at have very big flat spots. This is part of the explanation for what we found in Theorem 6.8.11.

The shift operators on ℓ_∞ are defined by

$$s_1(u_1, u_2, u_3, \ldots) = (u_2, u_3, u_4, \ldots) \text{ and } s_{j+1}(u) = s_1(s_j(u)). \tag{38}$$

Note that if $A \subset \mathbb{N}$ then $s_j(1_A) = 1_{A\ominus j}$ where $A \ominus j := \{t - j : t - j \geq 1, t \in A\}$.

Definition 6.9.1. $f \in \ell_\infty^\dagger$ *is a* ***Banach-Mazur limit*** *if*
1. $[u \geq 0] \Rightarrow [f(u) \geq 0]$,
2. $(\forall u \in \ell_\infty)(\forall j \in \mathbb{N})[f(u) = f(s_j(u))]$*, and*
3. $f(\mathbf{1}) = 1$ *(where* $\mathbf{1} = (1, 1, 1, \ldots)$*).*

Theorem 6.9.2. f *is a Banach-Mazur limit iff* $f(u) = \sum_{t=1}^{T} {}^*u_t\eta_t$ *where* $T \in {}^*\mathbb{N} \setminus \mathbb{N}$, $\eta_1, \ldots, \eta_T \geq 0$, $\sum_{t=1}^{T} \eta_t \simeq 1$ *and* $\sum_{t=1}^{T} |\eta_t - \eta_{t-1}| \simeq 0$ *(where* $\eta_0 := 0$*).*

Proof. To be inserted. □

Notice that the η's of Example 6.8.1 are uniform, hence have $|eta_1 - \eta_0| = \frac{1}{T}$ and for $T \geq t \geq 2$, $\eta_t - \eta_{t-1} = \frac{1}{T} - \frac{1}{T} = 0$.

The η's that appear in Theorem 6.9.2 are exactly the translation invariant probabilities.

Definition 6.9.3. η *is* ***translation invariant*** *if for all* $j \in \mathbb{N}$ *and all* $A \subset \mathbb{N}$, $\langle 1_{^*A}, \eta\rangle \simeq \langle s_j(1_{^*A}), \eta\rangle$*, that is, if* $\eta({}^*A) \simeq \eta({}^*(A \ominus j))$*.*

One direction of the following is easy, I am pretty sure that the other direction is true (from some standard results), but this still has the official status of a conjecture.

Theorem 6.9.4. *An sfp η is translation invariant iff $\sum_{t=1}^{T} |\eta_t - \eta_{t-1}| \simeq 0$.*

Proof. If $\sum_{t=1}^{T} |\eta_t - \eta_{t-1}| \simeq 0$, then $|\eta({}^*A) - \eta({}^*(A \ominus j)) \leq \sum_{t \in A} |\eta_t - \eta_{t-j}| \leq j \cdot \sum_t |\eta_t - \eta_{t-1}| \simeq 0$ so that η is translation invariant.

Now suppose that η is translation invariant. Let $A' = \{2k \in \mathbb{N} : \eta(2k-1) < \eta(2k)\}$ so that $|\eta(A') - \eta(A' \ominus 1)| = \sum_t |\eta_t - \eta_{t-1}|$. Some variant of over/underspill (and the Loeb space construction?) looking at the sets $A' \cap \{1, \ldots, N\}$ should complete the argument. □

Definition 6.9.5. *An sfp is of* ***remote shifted Cesaro type*** *if for some $j \in {}^*\mathbb{N}$, $\eta_t = \frac{1}{T}$ for all $t \in \{j+1, j+2, j+3, \ldots, j+T\}$.*

Theorem 6.9.6. *For $u \in \ell_\infty$, $\underline{L}_S(u) := \liminf_{T \to \infty} \left(\inf_{j \geq 1} \frac{1}{T} \sum_{t=1}^{T} u_{j+t} \right) = \inf_{\eta \in \mathcal{C}_s} {}^\circ\langle {}^*u, \eta \rangle$ where $\mathcal{C}_S$ is the set of remote sfp's of shifted Cesaro type.*

Proof. Let (T_k, j_k) be a sequence such that $\frac{1}{T_k} \sum_{t=1}^{T_k} u_{j_k+t} \to \underline{L}_S(u)$, let (T, j) be the equivalence class of the sequence $((T_1, j_1), (T_2, j_2), \ldots)$ and let $\eta_{T,j}$ be the uniform distribution on $\{j+1, j+2, \ldots, j+T\}$ so that ${}^\circ\langle {}^*u, \eta_{T,j} \rangle = \underline{L}_S(u)$. There cannot be a remote shifted Cesaro η with lower integral. □

The big result (to be covered about here) is that the set of remote sfp's of the shifted Cesaro type are the set of extreme points of the translation invariant probabilities. The implication: if we know something about what is optimal when the utility function is given by any one of these, we know what is true for all of the infinitely patient preferences.

6.10. **Some Hints.** For the problems on **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.

For the problems on **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 6.10.1 (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. $\square$

6.11. **Related Readings.** For infinitesimals and nonstandard analysis more generally, see [Lindstrøm, 1988]. For what we have used so far, see Ch. 11.1 - 11.2 in [Corbae et al., 2009].

For the game theory material, see also Ch. 8.3 - 8.4 in [Fudenberg and Tirole, 1991].

7. Moderately Elementary 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, 1987].

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

Definition 7.1.1. *A **random variable** is an element of $\mathbb{R}^\Omega$, typically denoted X or Y. If T is an interval, then $t \mapsto X_t$ from T to $\mathbb{R}^\Omega$ is a **stochastic process**.*

To put it slightly differently, the set of random variables is $\mathbb{R}^\Omega$, while a set of random variables indexed by a time set is a stochastic process.

7.1.1. *Some Examples.* The first example gives a random variable inducing the uniform distribution on $[0, 1]$. This is a good starting point for many reasons.

Example 7.1.1. $\Omega = \{1, \dots, n\}$, $P(A) = \frac{\#A}{n}$. *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.*

Comment: 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 all probabilities on all complete separable metric spaces.

Example 7.1.2 (A hyperfinite Brownian motion)**.** *Let $T = \{0, \frac{1}{N}, \frac{2}{N}, \dots, \frac{N-1}{N}, 1\}$ 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: $X(0, \omega) \equiv 0$, $X(1, \omega) = \frac{1}{\sqrt{n}}\omega_1$, $X(2, \omega) = \frac{1}{\sqrt{N}}(\omega_1 + \omega_2)$, $\dots$, $X(\frac{k}{N}, \omega) = \frac{1}{\sqrt{N}}\sum_{i=1}^k \omega_i$. This is a random walk model that moves through time in step sizes $dt := \frac{1}{N}$, and moves up and down $\pm\sqrt{1/N}$.*

Comment: 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 7.1.3 (A hyperfinite Poisson process). *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 $dt := \frac{1}{N}$ is the infinitesimal size of the incremental steps in the time set, and λ is limited and strictly positive. Define $Y(0,\omega') \equiv 0$, $Y(\frac{1}{N},\omega') = \omega'_1$, $Y(\frac{k}{N},\omega') = \sum_{i\leq k}\omega'_i$.*

Comment: 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 7.1.4. *We can glue the previous two examples as $\Omega \times \Omega'$ so that P and Q are independent. After doing that, we can define $Z(\frac{k}{N},(\omega,\omega')) = X(\frac{k}{N},\omega) + Y(\frac{k}{N},\omega')$.*

Comment: 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.

7.1.2. *Time Paths.* 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 a version of 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 when we work in $V(S)$ 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 called the Riesz representation theorem, and we will see a lovely proof of it using nonstandard techniques.

7.2. **Monitoring and Likelihood Ratios.** We are now going to compare a Poisson process model of monitoring with different intensities/arrival rates to a Brownian process model of monitoring with different drifts. In both cases, the information content of the signals is continuous. In the next subsection, we will consider Brownian models with different volatilities, and in these models, the information content is discontinuous.

For our *-finite models, we take $T = \{0, \frac{1}{N}, \frac{2}{N}, \ldots, \frac{N-1}{N}, 1\}$ with increments $dt = \frac{1}{N}$.

Example 7.2.1 (Poisson process monitoring). *If a high effort is being exerted, then in any time interval the probability of excellent news arriving is $\lambda_h dt$, if low effort it is $\lambda_l dt$ where $\lambda_h > \lambda_l$ and both are limited. We have two hypotheses: high effort all the time versus low effort all the time. Now, suppose that at time $\frac{k}{N}$ with $\circ\frac{k}{N} = r$ where r is non-zero, we have observed 0, 1, 2, or ... m instances of really excellent news. We are interested in the likelihood ratios, i.e. the factors by which we would update our prior probability that it was high effort or low effort. Let Y_k^λ count the number of instances of excellent news if λdt is the probability, that is $Y^\lambda = \sum_{i \leq k} X_i$ where the X_i are iid Bernoulli(λdt) random variables.*

$$P(Y_k^\lambda = 0) = \left(1 - \frac{\lambda}{N}\right)^k = \left(1 - \frac{\lambda\frac{k}{N}}{k}\right)^k \simeq e^{-\lambda r} \tag{39}$$

$$P(Y_k^\lambda = 1) = k\left(1 - \frac{\lambda}{N}\right)^{k-1}\left(\frac{\lambda}{N}\right)^1 \simeq (r\lambda)e^{-\lambda r} \tag{40}$$

$$P(Y_k^\lambda = 2) = \frac{k!}{(k-2)!2!}\left(1 - \frac{\lambda}{N}\right)^{k-2}\left(\frac{\lambda}{N}\right)^2 \simeq \frac{1}{2!}(r\lambda)^2 e^{-\lambda r} \tag{41}$$

$$P(Y_k^\lambda = m) = \frac{k!}{(k-m)!m!}\left(1 - \frac{\lambda}{N}\right)^{k-m}\left(\frac{\lambda}{N}\right)^m \simeq \frac{1}{m!}(r\lambda)^m e^{-\lambda r}. \tag{42}$$

This means that the likelihood ratios are $(\frac{\lambda_h}{\lambda_l})^{m+1}\frac{e^{-r\lambda_h}}{e^{-r\lambda_l}}$ if $Y_k = m$. This has two interesting properties: first, the more events happen by time r, the more information there is in the signal; second, for a given number of events, the larger is r, the less information there is. It also means that the information content of the signal is both limited and is a continuous function of time, that is, $\frac{k}{N} \simeq \frac{k'}{N}$ implies that the likelihood ratios are infinitely close to each other and imperfectly informative. To put it another way, the contrast between observing in nonstandard time and observing in standard time is infinitesimal, there is, for all practical purposes, more or less the same amount of information to be had observing the nonstandard details of the paths as there is in observing the standard parts of the path.

To get at Brownian monitoring as in Example 7.1.2 above, we change the distribution of the iid X_t so that at any non-infinitesimal $\tau := {}^\circ\frac{k}{N}$, the distribution is $\simeq N(\tau \cdot \mu, \tau)$, which is called a Brownian motion with a drift μ. The contrast between the nonstandard and the standard is again infinitesimal. Here, if we are trying to distinguish (say) a positive drift $\mu > 0$ from a 0 drift, the larger the number of upward jumps by time $\tau = {}^\circ\frac{k}{N}$ is a stronger signal for the positive drift, and the larger the drift or the longer the time interval, the more information we have.

Example 7.2.2 (Brownian monitoring: I). *Suppose that high effort leads to a positive drift $\mu > 0$ and low effort leads to a drift of 0. If the X_t are iid with $P(X_t = 1) = \frac{1}{2}(1+\gamma)$ and $P(X_t = -1) = \frac{1}{2}(1-\gamma)$, then $E\,X_t = \gamma$ so that the average slope of the process is $\gamma\sqrt{dt}/dt = \gamma/\sqrt{dt}$ so that $\gamma_\mu = \mu\sqrt{dt}$ delivers a Brownian drift of μ and a variance $k(\frac{1}{N} - r^2\frac{1}{N^2}) \simeq \tau := {}^\circ\frac{k}{N}$ at time point $\frac{k}{N}$.*

A sufficient statistic for the value of the process at $\frac{k}{N}$ *is* $\sum_{n \leq k} I_t$ *where* $I_n = 1_{X_n > 0}$. *This is counting the number of upward movements in k steps. To compare the hypothesis that the drift is* 0 *to the hypothesis that it is* μ, *we are comparing two binomial distributions,* $\text{Bin}(k, \frac{1}{2})$ *versus* $\text{Bin}\ (k, \frac{1}{2}(1+\gamma_\mu))$. *The difference in means is* $k\gamma_\mu = k\mu\frac{1}{\sqrt{N}}$, *the standard deviation is* $\frac{\sqrt{k}}{2}$, *and the ratio of the two is* $2\mu\sqrt{k/N} \simeq 2\mu\sqrt{\tau}$ *where* $\tau := {}^\circ\frac{k}{N}$.

Brownian monitoring in $V(S)$ has the property that at a time $r > 0$, we are comparing a $N(r\mu, r)$ to a $N(0, r)$, and if we observe that the process is equal to x at time r, then the likelihood ratio is

$$\frac{\frac{\kappa}{\kappa} e^{-\frac{(x-r)^2}{2r}}}{e^{-\frac{x^2}{2r}}} = e^{-\frac{1}{2r}[x^2-(x-r)^2]} = e^{\frac{1}{2r}[r(2x-r)]} = e^{x-\frac{r}{2}} \tag{43}$$

where κ is the constant that makes the densities integrate to 1. This is continuous in x and r, and is less than or greater than 1 as x is less than or greater than $\frac{r}{2}$. Given the symmetry of the Gaussian density, this makes sense — 0 is the mean if effort is low, r the mean if it is high, $\frac{r}{2}$ is half way between the two. Therefore, observations below $\frac{r}{2}$ tend to support the hypothesis that effort is low, while observations above $\frac{r}{2}$ tend to support the opposite hypothesis.

7.3. **Monitoring Volatilities.** The contrast with the case that changes in effort lead to changes in the Brownian motion volatility is extreme — here there is an infinite amount of information in any time interval $[0, \epsilon)$ if $\epsilon > 0$ is non-infinitesimal. We first go through how this works in $V(S)$, then give two different *-finite models with the same property. A crucial piece of knowledge from what follows is that the fourth moment of a standard normal distribution is $3\sigma^4$.

Example 7.3.1 (Brownian monitoring: II). *Working in* $V(S)$, *suppose that we have a usual Brownian motion, at every* $t \in [0, 1]$, $\xi_t \sim N(0, t)$, *or we have a Brownian motion with a different volatility,* $\xi_t^r \sim N(0, rt)$ *where* $r \neq 1$. *Divide the interval* $[0, \epsilon)$ *into n equally sized pieces, and compute the squares of the increments* $Y_1 := (\xi_{1\cdot\epsilon/n} - \xi_0)^2, Y_2 := (\xi_{2\cdot\epsilon/n} - \xi_{1\cdot\epsilon/n})^2, \ldots, Y_k := (\xi_{k\cdot\epsilon/n} - \xi_{(k-1)\cdot\epsilon/n})^2$ *as well as the squares of the increments* $Y_k^r := (\xi^r_{k\cdot\epsilon/n} - \xi^r_{(k-1)\cdot\epsilon/n})^2$, $k = 1, \ldots, n$. *Under the standard Brownian motion, the average of the* Y_k, $\frac{1}{n}\sum_{k \leq n} Y_k$, *is* $\frac{\epsilon^2}{n^2}$, *with the changed volatility, the average of the* Y_k^r *is* $\frac{r^2\epsilon^2}{n^2}$, *and the difference is* $(1-r^2)\epsilon^2\frac{1}{n^2}$. *The variance of* $\frac{1}{n}\sum_{k \leq n} Y_k$ *is* $\frac{1}{n^2} n 3\frac{\epsilon^4}{n^4} = 3\epsilon^2\frac{1}{n^5}$. *The ratio of the difference in means to the standard deviation is*

$$\frac{(r^2-1)\epsilon^2(1/n^2)}{\sqrt{3}\epsilon(1/n^{2.5})} = \epsilon\frac{r^2-1}{\sqrt{3}}\sqrt{n} \to_{n\uparrow\infty} \infty. \tag{44}$$

In the previous example, for fixed $\epsilon > 0$, all that matters is the number of times we subdivide the interval. So, if we take the continuity of the time set at which we can choose to make observations seriously enough, that is, if we believe that we can subdivide time arbitarily finely, we are certain of any change in drift by any $\epsilon > 0$.[4]

[4]For those who have seen the notation, the change of volatility event belongs to $\mathcal{F}_{0+}$ where $(\mathcal{F}_t)_{t\in[0,1]}$ is the Brownian filtration.

A different summary of the previous example is that if $W \in \Delta(C([0,1]))$ is the probability measure of the standard Brownian motion and W^r is the probability measure of the Brownian motion with a different drift, then there exists a set of continuous function, E, with the property that $W(E) = 1$ and $W^r(E) = 0$. See [Kakutani, 1948] for the grand-daddy of such results.

Example 7.3.2 (Brownian monitoring: III). *One way to change the volality of the* **-finite Brownian walk is to change the increments from* $\pm\sqrt{1/N}$ *to* $\pm\sqrt{r/N}$, $r \not\simeq 1$. *This means that after one observation, at time* $\frac{1}{N}$, *the change in volatility is revealed. Here the set of time paths changes from those with increments* $\pm\sqrt{1/N}$ *to those with increments* $\pm\sqrt{r/N}$, *and the disjointness of the supports is built in from the first step.*

An alternative, ***ternary*** *random walk does not build in a different set of time paths from the first step, but still has, as it must, the property that one learns of a change in volatility infinitely fast. Suppose that* $P(X^s_t = \pm\sqrt{1/N}) = \frac{1}{2}(1-s)$ *and* $P(X^s_t = 0) = s$, *while* $P(X^r_t = \pm\sqrt{1/N}) = \frac{1}{2}(1-\gamma)$ *and* $P(X^r_t = 0) = \gamma$. *The volatility of* $\xi^s(k/N,\cdot) := \sum_{t\le k} X^s_t$ *is given by* $\mathrm{Var}(\xi^s(k/N,\cdot)) \simeq (1-s)^\circ\frac{k}{N}$, *while the volatility of* $\xi^r(k/N,\cdot) := \sum_{t\le k} X^r_t$ *is given by* $\mathrm{Var}(\xi^r(k/N,\cdot)) \simeq (1-r)^\circ\frac{k}{N}$. *Supposing that* $r \not\simeq s$, *i.e. that the volatilities are not infinitesimally close to each other, the sufficient statistic for the difference between* $\xi^r(k/n,\cdot)$ *and* $\xi^s(k/N,\cdot)$ *is the number of size-*0 *increments by time* k. *This is revealed by any* $k \simeq \infty$, *and if* $\frac{k}{N} \simeq 0$ *at the same time, then we have learned the volatility difference before any standard time has elapsed.*

7.4. **A Brief Detour Through Queueing Theory.** In studying queues, the non-homogenous Poisson process has a central place. This involves the arrival rate, λ, varying with time. In our context, this would correspond to non-constant effort.

Example 7.4.1 (Non-constant effort). *Suppose that effort at each* $t \in T$ *yields a probability* $\lambda(t)dt$ *of there being excellent news in the t'th interval. We can show that* $\Pi_{i\le k}(1-\lambda(i)/N) \simeq e^{\frac{1}{N}\sum_{i\le k}\lambda(i)}$ *and, provided that* $t \mapsto \lambda(t)$ *is fairly well-behaved,* $\frac{1}{N}\sum_{i\le k}\lambda(i) \simeq \int_0^r {}^\circ\lambda(x)\,dx$.

7.5. **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, $E\,X := \sum_\omega X(\omega)P(\omega)$.

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

7.5.2. *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 $f : \mathbb{R} \to \mathbb{R}$,

$$f(\Sigma_\omega X(\omega)P(\omega)) \leq \Sigma_\omega f(X(\omega))P(\omega), \text{ equivalently} \tag{45}$$

$$f(E\,X) \leq E\,f(X), \tag{46}$$

provable from the definition of convexity and induction.

Lemma 7.5.1. *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 $\lim_{p\uparrow\infty} \|X\|_p = \|X\|_\infty$.

7.5.3. *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 7.5.2 (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

$$\Sigma_\omega |X(\omega)Y(\omega)|P(\omega) \leq \tfrac{1}{p}\Sigma_\omega\,|X(\omega)|^p\,P(\omega) + \tfrac{1}{q}\Sigma_\omega\,|Y(\omega)|^q\,P(\omega). \tag{47}$$

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

$$\log\left(\tfrac{1}{p}|X(\omega)|^p + \tfrac{1}{q}|Y(\omega)|^q\right) \geq \tfrac{1}{p}\log\left(|X(\omega)|^p\right) + \tfrac{1}{q}\log\left(|Y(\omega)|^q\right) \tag{48}$$

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

$$\log\left(\tfrac{1}{p}|X(\omega)|^p + \tfrac{1}{q}|Y(\omega)|^q\right) \geq \log\left(|X(\omega)Y(\omega)|\right). \tag{49}$$

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,

$$\tfrac{1}{p}\Sigma_\omega|X(\omega)|^p P(\omega) + \tfrac{1}{q}\Sigma_\omega|Y(\omega)|^q P(\omega) \geq \Sigma_\omega|X(\omega)Y(\omega)|P(\omega) \tag{50}$$

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 7.5.3. *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 7.5.4 (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. □

7.5.4. *The Markov/Chebyshev Inequality.* 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 **Markov's inequality**,

$$P(X \geq r) \leq \tfrac{1}{r}E\,X. \tag{51}$$

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

$$P(|X| \geq r) \leq \tfrac{1}{r^p}E\,|X|^p. \tag{52}$$

Often, the follow variant of this is called **Chebyshev's inequality**,

$$P(|X - E\,X| \geq r) \leq \tfrac{1}{r^2}\,\mathrm{Var}(X). \tag{53}$$

7.6. **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 $\mathrm{span}\,\{1_{A_k} : k = 1,\ldots K\}$ 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 7.6.1. *$\mathcal{A} \subset \mathbb{R}^\Omega$ is a **vector algebra** if for all $X, Y \in \mathcal{A}$ and all $\alpha, \beta \in \mathbb{R}$,*

a. $\alpha \cdot 1_\Omega \in \mathcal{A}$,

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

Atoms are, historically, the indissolubly small objects.

Definition 7.6.2. *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

$$X_\omega(\cdot) = \tfrac{Y(\cdot)-Y(\omega)}{Y(A)-Y(\omega)}. \tag{54}$$

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

$$R(\cdot) = \Pi_{\omega \notin A} X_\omega(\cdot). \tag{55}$$

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

Lemma 7.6.3. *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 7.6.4. *The mapping* $X \mapsto E\left(X|\mathcal{A}\right)$ *is orthogonal projection.*

7.6.1. *Algebras and Information.* Often, $\mathcal{A}$ will represent available information for a decision problem. Specifically, if one 'knows' $\mathcal{A} = \text{span}\left(\{1_{A_k} : k = 1, \ldots, K\}\right)$, then one solves the maximization problem

$$\max_{a_1,\ldots,a_K} \int \sum_k u(a_k, \omega) 1_{A_k}(\omega)\, dP(\omega). \tag{56}$$

The bridge-crossing logic is at work here, $a^* = (a_1^*, \ldots, a_K^*)$ solves this problem iff each a_k^* solves

$$\max_a \int u(a, \omega)\, dP(\omega|A_k). \tag{57}$$

The larger is $\mathcal{A}$, the smaller are the atoms, and the more functions $\omega \mapsto a(\omega) = (a_1(\omega), \ldots, a_K(\omega))$ one can choose.

7.7. **Adapted Stochastic Processes.** The time set, $T = \{t_0 < t_1 < t_2 < \cdots < t_N\}$ is a member of ${}^*\mathcal{P}_F(\mathbb{R})$ with the property that $(t_{n+1} - t_n) \simeq 0$ for $n = 0, \ldots, N-1$. We often take $t_1 = 0$ and $t_N = 1$, another frequent option has t_N unlimited.

Definition 7.7.1. *A* ***filtration*** *is an increasing mapping* $t \mapsto \mathcal{A}_t$ *from* T *to vector algebras in* $\mathbb{R}^\Omega$. *A stochastic process* $\xi : T \to \mathbb{R}^\Omega$ *is* ***adapted to the filtration*** $(\mathcal{A}_t)_{t\in T}$ *if for each* $t \in T$, $\xi(t, \cdot) \in \mathcal{A}_t$.

It is often useful to think of stochastic process as being 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 $\xi(\cdot, \omega)$. Given a stochastic process $\xi : T \times \Omega \to \mathbb{R}$, the **canonical filtration associated with** ξ has $\mathcal{A}_t$ defined as the smallest vector algebra of functions containing $\{\xi(s, \cdot) : s \leq t\}$.

8. Some Convergence Results in Probability Theory

8.1. **The Weak Law of Large Numbers (WLLN).** The weak law says that if we are averaging over a large number of independent/uncorrelated random variables, then it is "very likely" that the average is near the mean. The strong laws says that it is "very likely" that the average settles down at the mean.

8.1.1. *The easiest WLLN.* Here it is.

Theorem 8.1.1 (WLLN: I)**.** *If $\{\xi(n,\cdot) : n \in \{1,\ldots,N\}$ is an iid collection of random variables with mean 0 and variance 1, then for any non-infinitesimal $r > 0$, for any unlimited $n \leq N$, $P(\{\omega : \left|\frac{1}{n}\sum_{i\leq n}\xi(n,\omega)\right| < r\}) \simeq 1$.*

Proof. $\mathrm{Var}(\frac{1}{n}\sum_{i\leq n}\xi(n,\omega)) = \frac{1}{n^2}\cdot n = \frac{1}{n}$. By Chebyshev's inequality,

$$P(\{\omega : \left|\tfrac{1}{n}\textstyle\sum_{i\leq n}\xi(n,\omega)\right| \geq r\}) \leq \tfrac{1}{r^2 n} \simeq 0. \quad \square \tag{58}$$

Notation: let us replace $\xi(n,\cdot)$ with ξ_n and stop being so finicky about talking about the set of ω such that something is true.

Some easy extensions:

a. If r is an infinitesimal with $r^2 n$ infinite, then $P(\left|\frac{1}{n}\sum_{i\leq n}\xi_n\right| < r) \simeq 1$.
b. If the ξ_n are not iid but merely a collection of uncorrelated random variables with mean 0 and variance 1, the same conclusion holds.

8.1.2. *Triangular arrays.* The immediate impulse is to compare Theorem 8.1.1 to the following WLLN argument in $V(S)$: if $X_1, X_2, \ldots$ is an iid sequence of mean 0 variance 1 random variables, then $\mathrm{Var}(\frac{1}{n}\sum_{i\leq n}X_i) = \frac{1}{n}$ so that for any $r > 0$, $P(|\frac{1}{n}\sum_{i\leq n}X_i| \geq r) \leq \frac{1}{r^2 n} \to 0$. In fact, we have proved something much more general, formulated in terms of **triangular arrays** of random variables.

The unlimited N in $\{1,\ldots,N\}$ is $N = \langle N_1, N_2, N_3, \ldots\rangle$. This means that

$$\{1,\ldots,N\} = \langle\{1,\ldots,N_1\},\{1,\ldots,N_2\},\ldots\rangle. \tag{59}$$

This in turn means that $\{\xi_1,\ldots,\xi_N\}$ is the equivalence class of a sequence of sets of random variables arrayed in the following triangular fashion,

$$S_1 = \{\xi_{1,1},\ldots,\xi_{1,N_1}\} \tag{60}$$

$$S_2 = \{\xi_{2,1},\ldots,\xi_{2,N_1},\ldots,\xi_{2,N_2}\} \tag{61}$$

$$S_3 = \{\xi_{3,1},\ldots,\xi_{3,N_1},\ldots,\xi_{3,N_2},\ldots,\xi_{3,N_3}\} \tag{62}$$

$$\vdots$$

$$S_k = \{\xi_{k,1},\ldots,\xi_{k,N_1},\ldots,\xi_{k,N_2},\ldots,\xi_{k,N_3},\ldots,\xi_{k,N_k}\} \tag{63}$$

$$\vdots$$

The random variables in the set S_1 are defined on $(\Omega_1,\mathcal{F}_1,P_1)$ and have law $\mathcal{L}_1$, the random variables in the set S_2 are defined on $(\Omega_2,\mathcal{F}_2,P_2)$ and have law $\mathcal{L}_2$, the random variables in the set S_k are defined on $(\Omega_k,\mathcal{F}_k,P_k)$ and have law $\mathcal{L}_k$. The random variables $\xi_{k,n}$, $n = 1,\ldots,N_k$ are iid with law $\mathcal{L}_k$, but $\mathcal{L}_k$ need not equal $\mathcal{L}_{k'}$, and the random variables $\xi_{k,n}$ need not be independent of $\xi_{k',n'}$.

One can go a bit further in allowing the distributions to be different across the sets. The proof of the following is also directly from Chebyshev's inequality.

Theorem 8.1.2 (WLLN: II). *If* $\{\xi_n : n \in \{1, \ldots, N\}$ *is an uncorrelated set of random variables with* $\mathrm{Var}(\xi_n) = \sigma_n^2$*, then, defining* $s_n = \sum_{i \leq n} \sigma_i^2$*, for all* $n \leq N$ *with* s_n *is unlimited, there exists an infinitesimal* $r > 0$ *such that* $P(|\frac{1}{s_n} \sum_{i \leq n}(\xi_i - E\,\xi_i)| < r) \simeq 1$.

Defining $\mu_n = \sum_{i \leq n} E\,\xi_i$, this can be re-written as $P(|\frac{1}{s_n}((\sum_{i \leq n} \xi_i) - \mu_n)| < r) \simeq 1$.

8.2. **Almost Everywhere.** Defining $A_n(\omega) = \frac{1}{n} \sum_{i \leq n} \xi_i$. What the WLLN says is that, provided $\sum_{i \leq n} \sigma_i^2$ is unlimited, $P(A_n \simeq 0) \simeq 1$. This does **not** imply that the sequence $A_1(\omega), A_2(\omega), \ldots, A_N(\omega)$ is near converging to 0 for all of, or even most of, the $\omega \in \Omega$. However, this is what the Strong Law of Large Numbers (SLLN) says, and this is what we are building toward.

Let us say that a property $\mathbb{A}(\omega)$ holds **almost everywhere (a.e.)** if for every non-infinitesimal $\epsilon > 0$, there a set N with $P(N) < \epsilon$, and $\mathbb{A}(\omega)$ holds for all $\omega \in N^c$. (The capital "N" in the set N is a mnemonic for "null set.") The reason to formulate things this way is that we will often have the null set, N', of exceptions to the property we care about not being an internal set. However, it will contained in a larger internal set N with $P(N) \simeq 0$, or for any non-infinitesimal $\epsilon > 0$, it will be contained in an internal set N_ϵ with $P(N_\epsilon) < \epsilon$. The simplicity of the following should not lead you to underestimate it.

Theorem 8.2.1. *For any random variable* $X \in \mathbb{R}^\Omega$*, the following are equivalent:*
(a) $X \simeq 0$ *a.e.,*
(b) *for any non-infinitesimal* $r > 0$*,* $P(|X| \geq r) \simeq 0$*, and*
(c) *for some infinitesimal* $r > 0$*,* $P(|X| \geq r) \simeq 0$.

Proof. If $X \simeq 0$, then for any non-infinitesimal $\epsilon > 0$, there exists N with $P(N) < \epsilon$ such that $X(\omega) \simeq 0$ for all $\omega \notin N$. Therefore, for any non-infinitesimal $r > 0$, $\{|X| \geq r\} \subset N$. This being true for any non-infinitesimal $\epsilon > 0$, $P(|X| \geq r) \simeq 0$.

If for any non-infinitesimal $r > 0$, $P(|X| \geq r) \simeq 0$, then the set of r such that $P(|X| \geq r) \leq r$ contains arbitrarily small non-infinitesimal elements, hence contains an infinitesimal element.

If for some infinitesimal $r > 0$, $P(|X| \geq r) \simeq 0$, then $N^c := \{\omega : |X(\omega)| \leq r\}$ has $P(N^c) \simeq 1$. □

8.3. **Converging Almost Everywhere.** We are after a statement that the averages $A_n(\omega) := \frac{1}{n} \sum_{i \leq n} \xi_i$ converge to 0 almost everywhere when the ξ_i have mean 0. The following tells us that knowing about the probabilities of the maxima of collections of random variables is going to matter a huge amount.

Theorem 8.3.1. *The random variables* $X_1, \ldots, X_N$ *converge to* 0 *almost every iff for all non-infinitesimal* $r > 0$ *and all unlimited* $n \leq N$*,* $P(\max_{i \in \{n,N\}} |X_i| \geq r) \simeq 0$.

To understand what is going on back in $V(S)$, $N = \langle N_1, N_2, N_3, \ldots \rangle$ and $n = \langle n_1, n_2, n_3, \ldots \rangle$ with both sequences, N_k and n_k, going to ∞, though with $n_k \leq N_k$.

Going back to the triangular array, we are looking at, in the k'th row of (63), at the comparison of

$$X_{k,1},\ldots,X_{k,n_k} \text{ and} \tag{64}$$
$$X_{k,1},\ldots,X_{k,n_k},\ldots X_{k,N_k}. \tag{65}$$

One way to say that the random sequence X_{n_k} has "settled down" is to say that the stuff between X_{k,n_k} and X_{k,N_k} contributes nothing but infinitesimals.

Proof of Theorem 8.3.1. Define $M(n,r) = \{\omega : \max_{i\in\{n,N\}} |X_i(\omega)| \geq r\}$.

Suppose that $X_1,\ldots,X_N$ converges to 0 a.e. Pick non-infinitesimal $r > 0$ and $\epsilon > 0$. There exist N such that $P(N) < \epsilon$ and for all $\omega \in N^c$, $X_1(\omega),\ldots,X_N(\omega)$ converges to 0. By the definition of convergence, $X_n \simeq 0$ for all unlimited n. Therefore, $M(n,r) \subset N$ so that $P(M(n,r)) < \epsilon$. Since ϵ was arbitrary, $P(M(n,r)) \simeq 0$, i.e. $P(\max_{i\in\{n,N\}} |X_i| \geq r) \simeq 0$.

Now suppose that $P(M(n,r)) \simeq 0$ for all non-infinitesimal $r > 0$ and all unlimited $n \leq N$. Pick arbitrary non-infinitesimal $\epsilon > 0$. For $j \in {}^*\mathbb{N}$, define $n_j \in {}^*\mathbb{N}$ as the smallest integer such that

$$P(M(n_j,\tfrac{1}{j})) \leq \tfrac{\epsilon}{2^j}. \tag{66}$$

If j is limited, then overspill tells us that n_j must be finite. Let $N = \cup_{j\in{}^*\mathbb{N}} M(n_j,\frac{1}{j})$ so that $P(N) \leq \epsilon$. If $\omega \in N^c$, then for any unlimited $n \leq N$, $|X_n(\omega)| \leq \frac{1}{j}$ for any limited $j \in \mathbb{N}$, hence $|X_n(\omega)| \simeq 0$. □

8.4. **Weak Laws Versus Strong Laws.** The reason the argument for the weak law does not deliver the strong law is that we could have $P(|A_i| \geq r) \simeq 0$ for all unlimited i without having the probability that the maximum is above r being really small. The Poisson case is an example.

Example 8.4.1. *Let $X_1,\ldots,X_N$ be iid with $P(X_n = 1) = \lambda dt$ and $P(X_n = 0) = 1 - \lambda dt$ where $dt = \frac{1}{N} \simeq 0$. For each $i \in \{1,\ldots,N\}$ and any non-infinitesimal $r > 0$, $P(|X_i| \geq r) \simeq 0$, but $P(\max_{i\in\{N/2,N\}} |X_i| \geq 1)$ is not infinitesimal because it is the probability that a Poisson $\lambda/2$ is non-zero.*

8.4.1. *The Maximum of a Sum.* Kolmogorov's inequality talks about the probability that the maximum of a sum of independent random variables is large. This will make possible some uses of Theorem 8.3.1.

Theorem 8.4.1 (Kolmogorov's inequality)**.** *If $X_1,\ldots,X_n$ is a collection for independent mean 0 random variables with finite variance and $S_k := \sum_{i\leq k} X_i$, then for any $r > 0$,*

$$P\left(\max_{1\leq k\leq n} |S_k| \geq r\right) \leq \tfrac{1}{r^2}\,\mathrm{Var}(S_n). \tag{67}$$

This should be compared to Chebyshev's inequality, $P\left(|S_n| \geq r\right) \leq \frac{1}{r^2}\,\mathrm{Var}(S_n)$. We expect that $\max_{1\leq k\leq n} |S_k| \geq |S_n|$, so one version of what this is telling us is that if $\max_{1\leq k\leq n} |S_k|$ is large, then $|S_n|$ is also likely to be large.

Proof of Kolomogorov's inequality. Define τ as the minimum of the k such that $|S_k| \geq r$, so that $A_k := \{\tau = k\}$, $k = 1, \ldots, n$, partitions the event $\{\max_{1 \leq k \leq n} |S_k| \geq r\}$.

$$
\begin{aligned}
E\, S_n^2 &\geq \sum_k \int_{A_k} S_n^2 \, dP \\
&= \sum_k \int_{A_k} [S_k + (S_n - S_k)]^2 \, dP \\
&= \sum_k \int_{A_k} \left[S_k^2 + (S_n - S_k)^2 + 2S_k(S_n - S_k)\right] dP \\
&\geq \sum_k \int_{A_k} \left[S_k^2 + 2S_k(S_n - S_k)\right] dP.
\end{aligned} \tag{68}
$$

Note that A_k and S_k depend on $X_1, \ldots, X_k$, while $(S_n - S_k)$ depends on $X_{k+1}, \ldots, X_n$ and has mean 0. The two sets of random variables are independent. Therefore,

$$E\, 1_{A_k} S_k(S_n - S_k) = E\left(E\left(1_{A_k} S_k(S_n - S_k) \middle| X_1, \ldots, X_k\right)\right) = E\, 1_{A_k} S_k \cdot 0 = 0. \tag{69}$$

Using the fact that for every $\omega \in A_k$, $S_k^2(\omega) \geq r^2$, this yields

$$
\begin{aligned}
E\, S_n^2 &\geq \sum_k \int_{A_k} S_k^2 \, dP \geq r^2 \sum_k P(A_k) \\
&= r^2 P\left(\max_{1 \leq k \leq n} |S_k| \geq r\right). \quad \square
\end{aligned} \tag{70}
$$

8.4.2. *Sums with Random Signs.* We know that for any unlimited N, $\sum_{n=1}^{N} \frac{1}{n}$ is unlimited. Compare that result to the following consequence of Theorem 8.3.1.

Corollary 8.4.1.1. *If $R_1, \ldots, R_N$ is an iid sequence with $P(R_n = +1) = P(R_n = -1) = \frac{1}{2}$, then $Y_n := \sum_{i \leq n} R_i \frac{1}{i}$, $n = 1, \ldots, N$, converges a.e. to an a.e. limited random variable Y.*

Proof. Define $Y = Y_N$ so that $\mathrm{Var}(Y) \simeq \pi^2/6$. Having a finite variance, the probability that $|Y|$ is unlimited is infinitesimal, i.e. Y is a.e. limited. We want to show is that $X_n := Y_N - Y_{n-1}$ converges to 0 a.e. It is sufficient to show that $P(\max_{i \in \{n,N\}} |X_i| \geq r) \simeq 0$ for all unlimited n and non-infinitesimal $r > 0$. Each X_n is equal to $\sum_{i=n}^{N} R_i \frac{1}{i}$ so we are after the maximum of a set of sums of independent mean 0 random variables with finite variances. To make it fit exactly, we reverse the order of summation, define $Y_1' = R_{N-1} \frac{1}{N-1}$, $Y_2' = R_{N-2} \frac{1}{N-2}$, …, $Y_k' = R_{N-k} \frac{1}{N-k}$ and $S_k' = \sum_{i \leq k} Y_i'$ for $k = 1, \ldots (N - n)$. From Kolmogorov's inequality, $P\left(\max_{1 \leq k \leq n} |S_k'| \geq r\right) \leq \frac{1}{r^2} \mathrm{Var}(S_{N-n}')$. Now, $\mathrm{Var}(S_{N-n}') = \sum_{i=n}^{N} \frac{1}{i^2} \simeq 0$. $\square$

This gives us a huge number of examples of sequences x_n with $\sum |x_n|$ unlimited and $\sum x_n$ limited. One can show that Y has full support, which gives us the result that for any limited $x \in \mathbb{R}$, there is a sequence of $\pm$ signs r_n such that $\sum_{n \leq N} r_n \frac{1}{n} \to x$.

8.4.3. *A First Strong Law.*

Theorem 8.4.2 (Strong Law: I). *If $X_1, \ldots, X_N$ is a sequence of mean 0 random variables, N unlimited, and $\sum_{n=1}^{N} \frac{\text{Var}(X_n)}{n^2}$ then $A_N := \frac{1}{N}\sum_{n\leq N} X_n$ converges to 0 almost everywhere.*

Proof. We first show that $A_N \simeq 0$ almost everywhere, that is, $P(|A_N| > r) \simeq 0$: since N is unlimited, so is n' defined as the largest integer less than or equal to $\sqrt{N}$; this means that $\sum_{k=n'}^{N} \frac{\text{Var}(X_k)}{k^2} \simeq 0$; therefore A_N can be expressed as the sum of two independent random variables with infinitesimal variance, $A_N^{n'-1} := \frac{1}{N}\sum_{k=1}^{n'-1} X_k$ and $A_N^{n'+} := \frac{1}{N}\sum_{k=n}^{N} X_k$.

Pick unlimited $n \leq N$, and for $m \in \{n, \ldots N\}$, define $Y_m = A_N - A_m$ so that

$$Y_m = (\tfrac{1}{N} - \tfrac{1}{m})\textstyle\sum_{k=1}^{m-1} X_k + \frac{1}{N}\sum_{k=m}^{N} X_k. \tag{71}$$

The first term and the second term converge to 0. □

The usual proof uses the Borel-Cantelli lemmas to show that for all unlimited $n \leq N$ and non-infinitesimal $r > 0$, $P(\max_{m\in\{n,\ldots,N\}} |A_m| > r) \simeq 0$. They are useful also in understanding how much experimentation there must be in learning models so as to get past the Fudenberg and Levine self-confirming equilibria [Fudenberg and Levine, 1993].

8.5. **The Borel-Cantelli Lemmas.** In $V(S)$ the Borel-Cantelli lemma has the following form. If A_n is a sequence of events, define $[A_n \text{ i.o.}] = \cap_N \cup_{n\geq N} A_n$ and $[A_n \text{ a.a.}] = \cup_N \cap_{n\geq N} A_n$. Note that $([A_n \text{ i.o.}])^c = [A_n^c \text{ a.a.}]$ and $([A_n \text{ a.a.}])^c = [A_n^c \text{ i.o.}]$.

Lemma 8.5.1 (First Borel-Cantelli). *If A_n is a sequence of events and $\sum_{n\in\mathbb{N}} P(A_n) < \infty$, then $P([A_n \text{ i.o.}]) = 0$.*

Proof. $P([A_n \text{ i.o.}]) \leq \sum_{n\geq N} P(A_n) \downarrow 0$. □

Lemma 8.5.2 (Second Borel-Cantelli). *If A_n is a sequence of independent events and $\sum_{n\in\mathbb{N}} P(A_n) = \infty$, then $P([A_n \text{ i.o.}]) = 1$.*

Proof. $([A_n \text{ i.o.}])^c = [A_n^c \text{ a.a.}] \subset \cap_{n\geq N} A_n^c$ so it is sufficient to show that $P(\cap_{n\geq N} A_n^c) = 0$ for any N. Using $(1 - x) \leq e^{-x}$,

$$P\left(\cap_{k=N}^{N+j} A_k^c\right) = \Pi_{k=N}^{N+j}(1 - P(A_k^c)) \leq \Pi_{k=N}^{N+j} e^{-P(A_k)} = e^{-\sum_{k=n}^{N+j} P(A_k)}. \tag{72}$$

Since $\sum_{k=n}^{N+j} P(A_k) \to \infty$ as $j \uparrow$, the right-hand side goes to 0.

Here are the nonstandard versions of these results. They are key ingredients in the proofs of some of the forms of strong law of large numbers. They're also pretty cool, and useful too.

The weak law told us that $A_n \simeq 0$ almost everywhere. By contrast, the strong law tells us that A_n is near convergent to 0 almost everywhere.

8.6. **Limited Fluctuation.** Because $\mathbb{R}$ is complete in the usual metric, the convergent sequences in $V(S)$ are the Cauchy sequences. Cauchy sequences are the ones that do not vary by more than any $\epsilon \in \mathbb{R}_{++}$ more than finitely many times.

Suppose that $\xi : [0, \infty) \to \mathbb{R}$ in $V(S)$. Here is a Cauchy-style reformulation of ξ converging: ξ has k ϵ-flucations if there exist time points $t_0, \ldots, t_k \in [0, \infty)$ such that

$|\xi(t_1) - \xi(t_0)| \geq \epsilon$, $|\xi(t_2) - \xi(t_1)| \geq \epsilon$, …, $|\xi(t_k) - \xi(t_{k-1})| \geq \epsilon$; ξ converges if for all $\epsilon > 0$, it has only finitely many ϵ-fluctations.

Near converging and having a limited number of ϵ-fluctuations are different for internal $\xi : T \to {}^*\mathbb{R}$.

Definition 8.6.1. *An internal $\xi : T \to {}^*\mathbb{R}$ is of* ***limited fluctuation*** *if for all non-infinitesimal $\epsilon > 0$, ξ does not have an unlimited number of ϵ-fluctations.*

The set of limited fluctuation ξ's is not an internal set. We are going to want to assign a probability to the set of ω's for which $n \mapsto \xi(n, \omega)$ has limited fluctuation.

Example 8.6.1. *For a sequence $n \mapsto \xi(n)$, $n \in \{1, \ldots, N\}$, N unlimited, and for an unlimited even $m < N$, let $\xi(n) = 0$ for $n \leq m/2$, $\xi(n) = 1$ for $m/2 < n \leq N$. On the initial interval $\{1, \ldots, m/2\}$, $n \mapsto \xi(n)$ near converges to 0, but it does not near converge on $\{1, \ldots, N\}$.*

Initial intervals converging to 0 is the general pattern for sequences having limited fluctuations. To prove it, we are going to map from $V({}^*S)$ to $\mathbb{N}$: let $A(\cdot)$ be a statement, internal or external, about $n \in \mathbb{N}$; if $\{n \in \mathbb{N} : A(n)\} \neq \emptyset$, then it has a least element. The following is about the set of limited fluctuation time paths, that is, it is about a set of time paths that is not internal. That's okay, we're only working with one at a time.

Theorem 8.6.2. *If $\xi : T \to \mathbb{R}$ is of limited fluctuation, then for some unlimited m, $\xi(t_0), \xi(t_1), \ldots, \xi(t_m)$ converges.*

Proof. The essential intution is that looking at the limited parts of the index set, the sequence has to be Cauchy. The actual argument is a bit trickier. □

8.7. **Versions of Regularity for Time Paths.**

8.7.1. *Time Path Properties.*

8.7.2. *Asymptotic similarities.* One way to talk about asymptotic similarity of infinite time paths, $x_t, y_t \geq 0$ in $V(S)$ is to say that $x = O(y)$, read as "x is big Oh of y," if $\limsup_{t\uparrow\infty} \frac{x_t}{y_t} < \infty$, $x = o(y)$, read as "x is little Oh of y," if $\limsup_{t\uparrow\infty} \frac{x_t}{y_t} = 0$, and "$x$ is asymptotic to y" if $\lim_{t\uparrow\infty} \frac{x_t}{y_t} = 1$.

Definition 8.7.1. *For $x, y \in {}^*\mathbb{R}$, x* ***is asymptotic to*** *y if $\frac{x}{y} \simeq 1$.*

The "big O" and "little o" relations are $\frac{x}{y}$ being limited and being infinitesimal.

The following two lemmas and example tell us that an infinitesimal percentage error in an unlimited quantity can be infinitely large, but that an infinitesimal percentage error of a limited quantity is not large.

Lemma 8.7.2. *For limited $x, y \in {}^*\mathbb{R}$, x is asymptotic to y iff $x \simeq y$.*

Example 8.7.1. *If m is an unlimited integer, $x = m! + m$, and $y = m!$, then x is asymptotic to y and $|x - y|$ is unlimited.*

Lemma 8.7.3. *For all $n \in {}^*\mathbb{N}$, if for all $i \in \{1, \ldots n\}$, $x_i > 0$, $y_i > 0$ and x_i is asymptotic to y, then $\sum_{i \leq n} x_i$ is asymptotic to $\sum_{i \leq n} y_i$.*

Proof. For arbitrary $\epsilon \in \mathbb{R}_{++}$, $x_i \in [(1-\epsilon)y_i, (1+\epsilon)y_i]$, summing over i yields $\sum_{i\le n} x_i \in [(1-\epsilon)\sum_{i\le n} y_i, (1+\epsilon)\sum_{i\le n} y_i]$. □

9. Time Paths, Near Continuity, Integrals, and Control

Throughout, we are going to work with a time set $T = \{t_1, \dots, t_N\} \subset {}^*\mathbb{R}$ that is a **near interval**, that is, one that satisfies $d_H(T, {}^*[a,b]) \simeq 0$ for a limited interval $[a,b]$ where $d_H(\cdot,\cdot)$ is the Hausdorff metric.[5] This entail T having infinitesimal increments, which in turn entails N being unlimited. Further, we are going to assume that ${}^\circ t_1 = a$ and ${}^\circ t_N = b$.

To save on levels of subscripts in notation, a time path $(\xi(t_1,\omega), \xi(t_2,\omega), \dots, \xi(t_N,\omega))$ for $\omega \in \Omega$ will be denoted $(\xi(1,\omega), \xi(2,\omega), \dots, \xi(N,\omega))$, and even $(\xi_1, \dots, \xi_N)$.

9.1. Near Convergence.

A first useful properties of a time paths is that it "converges."

Definition 9.1.1. *For unlimited N, a sequence $n \mapsto \xi_n$, $n \in \{1, \dots, N\}$ is **near convergent** if for some $x \in {}^*\mathbb{R}$, $\xi_n \simeq x$ for all unlimited n.*

This means that from n to N, there is not very much movement in the sequence. In more detail, since the set $\{n, n+1, \dots, N\}$ is *-finite, there is an n° that solves $\max\{|\xi_m - x| : m \in \{n, \dots, N\}\}$, and by assumption, $|\xi_{n^\circ} - x| \simeq 0$.

Lemma 9.1.2. *If a sequence $n \mapsto \xi_n$ is both near convergent to an $x \in {}^*\mathbb{R}$ and limited for limited n, then x is nearstandard.*

Proof. If $n \mapsto \xi(n)$, $n \in \{1, \dots, N\}$, is near convergent to x, then $\{n \in \{1, \dots, N\} : |\xi(n) - x| < 1$ contains arbitrarily small infinite integers. Since this set is internal, it must contain finite integers. For any of those finite integers, $\xi(n)$ is limited, hence x is limited. □

Taking $\xi_n \equiv x$ for some unlimited x tells you that you need the assumption that ξ_n is limited for limited n in the previous.

Lemma 9.1.3. *If ξ_n near converges to an unlimited x, then it is unlimited for some finite n.*

Proof. $\{n \in \{1, \dots, N : (\forall m \in \{n, \dots, N\})|\xi_m - x| < 1\}$ contains arbitrarily small infinite integers, hence contains a finite integer. □

The following non-convergent example will help the understanding of triangular arrays of random variables.

Example 9.1.1. *Let $N = \langle 1, 2, 4, 8, 16, \dots\rangle$, for $k \in \mathbb{N}$, let define $\xi_k(n) = \min\{k^n, n^n\}$ for $n = 1, \dots, 2^k$, and let $n \mapsto \xi(n)$ be $\langle \xi_1(\cdot), \xi_2(\cdot), \xi_3(\cdot), \dots\rangle$. To describe this class of examples without using equivalence classes, let k be an infinite integer, set $N = 2^k$, and define $\xi(n) = \min\{k^n, n^n\}$ for $n \in \{1, \dots, N\}$.*

[5] Let $\mathcal{K}(\mathbb{R})$ denote the non-empty compact subsets of $\mathbb{R}$, for $K \in \mathcal{K}(\mathbb{R})$ and $\epsilon > 0$, define the ϵ-ball around K as $K^\epsilon = \cup_{x\in K} B_\epsilon(x)$, and define the Hausdorff metric by $d_H(K_1, K_2) = \inf\{\epsilon > 0 : K_1 \subset K_2^\epsilon, K_2 \subset K_1^\epsilon\}$.

9.2. **Near Continuity.** Convergence and continuity go together like beer and tequila.

Definition 9.2.1. *A function* $\xi : T \to {}^*\mathbb{R}$ *is* ***near continuous at*** t *if* $[s \simeq t] \Rightarrow [\xi(s) \simeq \xi(t)]$*, and it is* ***near continuous on*** T *if it is near continuous at every* $t \in T$.

After rearranging the indexing a bit, this is a statement about the behavior of $\xi(\cdot)$ along *-finite sequences in T that converge to t. We will use this observation in the proof of the following.

Lemma 9.2.2. *If* $\xi(\cdot)$ *is limited for some* $t \in T$ *and is near continuous, then it is uniformly bounded by a limited number, that is, there exists a limited* B *such that* $|\xi(t)| \leq B$ *for all* $t \in T$.

Proof. Let t_K solve $\max_{t \in T} |\xi(t)|$ and suppose that $\xi(t_K)$ is unlimited, and let $\xi(t_k)$ be limited. If $t_k < t_K$, consider the sequence $\xi(t_k), \xi(t_{k+1}), \ldots, \xi(t_K)$, if $t_k > t_K$, consider the sequence $\xi(t_k), \xi(t_{k-1}), \ldots, \xi(t_K)$. Near continuity and Lemma 9.1.2 tell us that $\xi(t_{k+m})$ is unlimited for some finite m, contradicting at t_k. □

We want to know that the nonstandard function $\xi : T \to \mathbb{R}$ is behaving like a sensible standard function on $[a, b]$. The problem is that every $\tau \in [a, b]$ corresponds to the many many $t \in T$ that have $^\circ t = \tau$. We are now going to show that a near continuous ξ that is limited at some $t T$ has a graph infinitesimally close to a standard $f \in C([a, b])$. Note that the boundedness guaranteed in Lemma 9.2.2 is necessary for this since any standard continuous function on $[a, b]$ is bounded.

Theorem 9.2.3. *If* T *is a near standard interval,* $\xi : T \to {}^*\mathbb{R}$ *is nearly continuous on* T*, and* $\xi(t)$ *is limited for some* $t \in T$*, then the standard part of the graph of* ξ *is the graph of a continuous function.*

Another way to put this is that we can define a function f in $V(S)$ by defining it as $f(t) = {}^\circ\xi(t')$ for any $t' \simeq t$, equivalently, $f(t) = \mathrm{st}\,(\xi(\mathrm{st}^{-1}(t)))$ where $\mathrm{st} : {}^*\mathbb{R} \to \mathbb{R}$ is the standard part mapping.

Proof. $gr(\xi)$, the graph of ξ, is an internal subset of $^*[a, b] \times {}^*[r, s]$, hence $^\circ gr(\xi)$ is a closed set, that is, it is the graph of an upper hemicontinuous, non-empty valued correspondence, call it Ξ. All that is needed for it to be the graph of a continuous function is that the correspondence be single-valued. Suppose not, i.e. suppose that for some $\tau \in [a, b]$, $x \neq x' \in \Xi(\tau)$. Pick a non-infinitesimal $\epsilon > 0$ such that $|x - x'| > \epsilon$. Let

$$D_\epsilon = \{\delta \in {}^*\mathbb{R}_{++} : (\forall t' \in T)[\,[|t - t'| < \delta] \Rightarrow [|\xi(t') - \xi(t)| < \epsilon]\,]\}. \tag{73}$$

D_ϵ contains arbitarily large infinitesimals, hence contains a non-infinitesimal $\delta > 0$. This means that if $x \in \Xi(\tau)$ then $x' \notin \Xi(\tau)$ because for every $\tau' \in B_\delta(\tau)$, $\Xi(\tau') \subset B_\epsilon(x)$. □

Definition 9.2.4. *If* (M, d) *is a metric space in* $V(S)$ *and* E *is an interval subset of* $^*M$*, then the* ***standard part of*** E *is* $\mathrm{st}\,(E) := \{x \in M : d(x, E) \simeq 0\}$.

Since we identify functions with their graphs, and their graphs are sets, we can write the conclusion of Theorem 9.2.3 as $\mathrm{st}\,(\xi) \in C([a, b])$.

9.3. **Paths and Integrals I.** Recall the first integrals you saw, from calculus class. A typical example is a continuous $f : [0,1] \to \mathbb{R}$, and we are interested in properties of $F(x) := \int_0^x f(t)\,dt$, $x \in [0,1]$. Remember that for any subdivision $0 = t_0 < t_1 < \cdots < t_N = 1$,

$$F_- := \sum_{n=1}^N \min_{x\in[t_{n-1},t_n]}(t_n - t_{n-1}) \leq \int_0^1 f(t)\,dt \leq F_+ := \sum_{n=1}^N \max_{x\in[t_{n-1},t_n]}(t_n - t_{n-1}), \tag{74}$$

that the upper and lower bounds converge to each other as the maximal length of the subdivisions goes to 0, and that the limit is independent of the sequence of subdivisions. This means that for any $f \in C([a,b])$, $\int_a^b f(t)\,dt \simeq \sum_{t\in T} {}^*f(t)\,dt$ where the symbol "dt" is being used in two different, but very similar ways in the integral and the *-finite sum.

Lemma 9.3.1. *If T is a near standard interval with endpoints a, b, $\xi : T \to {}^*\mathbb{R}$ is near continuous and limited on T, and $f = \text{st}\,(\xi)$, then*

$$\sum_t \xi(t)\,dt \simeq \int_a^b f(t)\,dt. \tag{75}$$

Proof. From the discussion of the first integrals seen in calculus classes, $\int_a^b f(t)\,dt \simeq \sum_{t\in t} {}^*f(t)\,dt$. Now, $|\sum_t \xi(t)\,dt - \sum_t {}^*f(t)\,dt| \leq \sum_t |\xi(t) - {}^*f(t)|\,dt$, and this is turn less than or equal to $N \cdot \max_{t\in T} |\xi(t) - {}^*f(t)|\,\frac{1}{N}$. Since T is *-finite, the maximum is achieved, and is $\simeq 0$. □

We will later be interested in two generalizations of Lemma 9.3.1.

1. If f is a piecewise-continuous function, just replace $\int_0^1 f(t)\,dt$ with $\sum_{i=0}^I \int_{\tau_i}^{\tau_{i+1}} f(t)\,dt$ where $0 = \tau_0 < \tau_1 < \cdots < \tau_I = 1$ and $\tau_1, \ldots, \tau_{I-1}$ are the discontinuity points of f.
2. If f is a measurable function, then a result called Lusin's theorem tells us that it is "nearly" a continuous function. This will allow us to show that integrals of measurable functions are infinitesimally close to *-finite sums over near intervals for the case that $t \mapsto f(t)$ is a measurable function. This will be useful when we consider $f(t) = \dot{x}(t)$ for a calculus of variations or a control problem, and in these cases, we will have $x(t) = x_0 + \int_0^t \dot{x}(s)\,ds$.

9.4. **A First Control Problem.** Here is one of the simplest control problems we will ever see in $V(S)$,

$$\min \int_0^1 [c_1\dot{x}^2(t) + c_2 x(t)]\,dt \text{ s.t. } x(0) = 0, x(1) \geq B, \dot{x}(t) \geq 0. \tag{76}$$

The idea is that, starting with none on hand, we need to produce a total amount B of a good by time 1. There is a storage cost, c_2, per unit stored for a unit of time, and producing at a rate r, i.e. having $x'(t) = dx/dt = \dot{x} = r$, incurs costs at a rate $c_1(x'(t))^2$. The tradeoff between producing fast productino rate and storage costs leads us to believe that the solution must involve starting production at a low level at some point in the interval and increasing the rate at which we produce as we near the end of the interval.

9.4.1. *A Near Interval Formulation.* Let us turn to the near interval formulation of the problem. We replace $[0,1]$ by a near interval T with increments dt, and to make life simpler we suppose that the increments have equal size, $dt \simeq \frac{1}{N}$ for some unlimited N. Now $x'(t)$ is the action, a_t, chosen at t, $a_t = \frac{x(t+dt)-x(t)}{dt}$, that is, the discrete slope of the amount on stock over the interval of time between t and $t+dt$. This means that if we choose actions $a_0, a_1, \ldots, a_{N-1}$, then by any $t \in T$, $x(t) = \sum_{s<t} a_s\, ds$. The problem is replaced by

$$\min_{a_0,a_1,\ldots,a_{N-1}} \sum_t \left[c_1 a_t^2 + c_2 \sum_{s<t} a_s\, ds\right] dt \text{ s.t. } \sum_{s<T} a_s\, ds = B,\ a_t \geq 0. \tag{77}$$

9.4.2. *Solving the Near Interval Formulation.* Writing the Lagrangean, we have

$$L(a_0,\ldots,a_{N-1};\lambda) = \sum_t \left[c_1 a_t^2 + c_2 \sum_{s<t} a_s\, ds\right] dt + \lambda(B - \sum_{s<T} a_s\, ds). \tag{78}$$

With the non-negativity constraints, the Kuhn-Tucker conditions are $\partial L/\partial a_t \leq 0$, $a_t \geq 0$, $a_t \cdot (\partial L/\partial a_t) = 0$. When the optimal a_t^* are positive, we have

$$\frac{\partial L}{\partial a_t} = \left[2c_1 a_t + c_2 \sum_{k=t}^{N-1} 1\, ds\right] dt - \lambda\, dt = 0. \tag{79}$$

Now, the dt's are all equal to $\frac{1}{N}$, and being a common factor, we can take them out. Further, $\sum_{k=t}^{N-1} 1\, ds \simeq (1-t)$. Putting these together, for the positive a_t^*'s, we have

$$a_t^* = \frac{\lambda - c_2(1-t)}{2c_1}. \tag{80}$$

This means that the slope is an affine function of t with slope $\frac{c_2}{2c_1}$: the larger is the storage cost, c_2, relative to c_1, the steeper the slope and the less time will be spent producing; the larger is the production cost, c_1, relative to c_2, the lower the slope, and the more time will be spent producing. All that is left is to clear up the remaining details.

To find the value of λ, we substitute back into the constraint $\sum_{s<T} a_s^*\, ds = B$, and use our result about integrals, Lemma 9.3.1, to find that we determine λ by solving

$$\int_0^1 \frac{\lambda - c_2(1-t)}{2c_1}\, dt = B, \tag{81}$$

which yields $\lambda^* = B \cdot 2c_1 + \frac{1}{2}c_2$. Finding the $\underline{t}$ at which we start producing involves setting $\lambda^* - c_2(1-t) = 0$, i.e. $\underline{t} = \left[\frac{1}{2} - \frac{2Bc_1}{c_2}\right]$ as long as this is non-negative, and $\underline{t} = 0$ otherwise, which agrees with the Kuhn-Tucker non-negativity versions of (80).

9.4.3. *Checking that We've Solved the Original Problem.* Now we check that we have indeed solved the original problem in $V(S)$, the one given in (1). The standard part of the near continuous function $t \mapsto a_t^*$ is 0 until $\max\{0, \underline{t}\}$ and it increases linearly with slope $\frac{c_2}{2c_1}$ until $t = 1$. Suppose that there exists a piecewise continuous alternative strategy $b(t) = \dot{x}(t)$ that yields more than $\text{st}\,(a^*(\cdot))$ in (1). This means that we can do better by a non-infinitesimal. Consider the strategy ${}^*b : T \to {}^*\mathbb{R}_+$, it must give within an infinitesimal of the strategy b in (1), which means that ${}^*b$ must beat a^* by a non-infinitesimal, a contradiction.

9.5. **A Control Problem Without a Solution.** Everything worked very smoothly in the first problem. Here is the simplest example that I know of about what can go wrong.

Consider the problem in $V(S)$,

$$\max \int_0^1 [\dot{x}^2(t) - x^2(t)]\, dt \text{ s.t. } -1 \leq \dot{x}(t) \leq +1 \tag{82}$$

where the maximum is taken over piecewise continuous functions $t \mapsto \dot{x}(t)$. The first term in the integrand tells us that we want to be moving as fast as possible, the second term tells us that we want to minimize our displacement. These have a rather contradictory feel to them. Let us examine just how contradictory.

9.5.1. *Nonexistence.* Divide the interval $[0, 1]$ into N equally sized sub-intervals, $N \in \mathbb{N}$, and consider the path that over each interval $[\frac{k}{N}, \frac{k+1}{N}]$ has $\dot{x} = +1$ for the first half and $\dot{x} = -1$ for the second half. This means that $x(t)$ goes up at slope $+1$ over the first half of each interval and down with slope -1 over the second half of each interval. An $N \uparrow$, the value to this path in (82) converges up to 1. However, the value 1 cannot be achieved by any path — *that* requires that $\dot{x}(t)$ alway be either ± 1 and $x(t)$ always be 0, contradictory requirements.

9.5.2. *A Near Interval Formulation.* Replace $[0, 1]$ with a near interval $T = \{0, \frac{1}{N}, \ldots, \frac{N-1}{N}, 1\}$ with N unlimited and even. Reformulate (82) as

$$\max_{a_0, a_1, \ldots, a_{N-1}} \sum_t \left[a_t^2 - \left(\sum_{s<t} a_s\, ds \right)^2 \right] dt \text{ s.t. } -1 \leq a_t \leq +1. \tag{83}$$

Notice that there is a pair of multipliers for each $t \in T$, one for the constraint $-1 \leq a_t$ and one for the constraint $a_t \leq +1$. Only one of each of these constraints can be binding at any point in time. Often the time path of the multipliers is very informative about when and where constraints are most strongly pinching the solution. Here, there is so much symmetry that the pattern of the multipliers looks like the pattern we will see in the solutions and has no further information.

One of the two solutions to this is $a^*_{k/N} = +1$ for the even k and $a^*_{k/N} = -1$ for the odd k (the other solution reverses the signs). This gives a utility of $1 - \frac{1}{2}dt^2 \simeq 1$. We see a continuation of the pattern of approach the supremum value of 1 — since $dt = \frac{1}{N}$, larger N yields smaller dt, yielding a higher value. Thus, the near interval formulation has a solution, it gives a value $\simeq 1$.

9.5.3. *Trying to Take the Near Interval Solution Back to $V(S)$.* The optimal path $t \mapsto a^*_t$ is *not* near continuous, it move up or down by 2 between each t and $t + dt$. This is a phenomenom known as **chattering**. Not only is there not any continuous function that behaves like this, there is not any *measurable* function.

To see why, look up Lebesgue's density theorem, it tells us that for any measurable $A \subset [0, 1]$, there is an $A' \subset A$ such $Unif(A \setminus A') = 0$ and for each $x \in A'$,

$$\lim_{\epsilon \downarrow 0} \frac{Unif(A \cap (x - \epsilon, x + \epsilon))}{2\epsilon} = 1. \tag{84}$$

If $A = [a, b]$, then $A' = (a, b)$, we just get rid of the end-points. The amazing part of Lebesgue's result is that this simple intuition of getting rid of end points works for all measurable sets. In particular, this means that for each $x \in A'$, the derivative of the function $H(x) = Unif(A\cap[0,x])$ is equal to 1. Applying Lebesgue's density theorem to $B := A^c$, for $x \in B'$, the derivative of $H(x)$ is equal to 0. Since $Unif(B)+Unif(B') = 1$, this means that the derivative of H is, for Lebesgue almost every $x \in [0,1]$, either equal to 0 or equal to 1.

Now, if there is a measurable function representing $t \mapsto a^*_t$, then we can partition $[0,1]$ into a set A on which $\dot{x}(t) = +1$ and $B := A^c$ on which $\dot{x}(t) = -1$. However, for every non-infinitesimal $x \in [0,1]$, the proportion of the $t \in T$ with $t < x$ and $a^*_t = +1$ is, up to an infinitesimal, equal to $\frac{1}{2}$. This means that for our measurable function we would have to have $Unif(A \cap [0,x]) = Unif(B \cap [0,x]) = \frac{1}{2}x$ for each $x \in (0,1]$. Lebesgue's density theorem tells us that this cannot happen, we must have the derivative equal to 1 or 0 almost everywhere rather than equal to $\frac{1}{2}$ everywhere.

9.5.4. *Another Represenation.* If we look at subsets $E := (a,b] \times (c,d]$ in $[0,1] \times [-1,+1]$, we could ask what proportion of the time is the path (t, a^*_t) in the set ${}^*E$? The answer gives a probability distribution that is half times the uniform distribution on $[0,1] \times \{+1\}$ and half times the uniform distribution on $[0,1] \times \{-1\}$. This is one version of what is often called a **Young measure**. We are not going to spend a lot of time on these right now, we will come back to them in some control problems, and a little bit more extensively while doing continuous time games.

9.6. **The Euler-Lagrange Necessary Conditions.** We are going to be studying two kinds of problems,

$$\max_{x(\cdot)} \int_0^1 U(t, x(t), \dot{x}(t))\, dt \text{ s.t. } x(0) = b_0,\ x(1) = b_1, \text{ and} \tag{85}$$

$$\max_{a(\cdot)} \int_0^1 U(t, x(t), u(t))\, dt \text{ s.t. } x(0) = b_0,\ x(1) = b_1,\ (\forall t \in [0,1])[a(t) \in A], \\ \text{and } (\forall t \in [0,1])[\dot{x}(t) = f(t, x(t), a(t))], \tag{86}$$

where the $x(\cdot)$ should be picked with $\dot{x}(\cdot)$ piecewise continuous, and the $a(\cdot)$ should also be piecewise continuous. Since $x(t) = x_0 + \int_0^t \dot{x}(s)\, ds$, we could reformulate the first problem as a maximization over piecewise continuous $a(\cdot)$'s with $\dot{x}(t) = a(t)$. To put it another way, taking $f(t, x(t), a(t)) \equiv a(t)$ and $A = \mathbb{R}$, the first problem becomes a special case of the second problem.

We will, mostly, study these problems by studying their near interval formulations and then either take the standard part of the optimal path, or work with the path on the near interval. The necessary conditions for an optimum for the first class of problems are called the **Euler-Lagrange conditions**, or sometimes the **Euler equation**. We will set up the second class of problems with an infinite set of constraints and a corresponding infinite set of multipliers, one on two for each t in a near interval T depending on whether the $a(t) \in A$ constraint is binding. When we use Lagrangean multipliers to set up a constrained problem, the necessary conditions are those of an unconstrained problem, so the Euler-Lagrange conditions will be useful there too.

9.6.1. *The Near Interval Formulations.* Replace $[0,1]$ with $T = \{\frac{k}{N} : k = 0, 1, \ldots, N\}$, $N \simeq \infty$. We will adopt the following notations for the first class of problems: dt or

ds denotes $\frac{1}{N}$; x_k denotes $x(\frac{k}{N})$; a_k denotes $a(\frac{k}{N})$; for $k = 1, 2, \ldots, N$, $x_k = \sum_{j<k} a_k dt$; and $a_k = \frac{x_{k+1}-x_k}{dt}$. This uses x_k as the state of the system at the beginning of the time interval $[\frac{k}{N}, \frac{k+1}{N})$, and has the action, a_k, determine the value of x_{k+1}, with a_k by being the slope of the line segment joining $(\frac{k}{N}, x_k)$ and $(\frac{k+1}{N}, x_{k+1})$, $k \leq (N-1)$.

With these notations, the first of the problems can be formulated either in terms of the optimal path of states or in terms of the optimal path of actions, that is, either as

$$\max_{x_k:k=1,\ldots,N-1} \sum_{k=0}^{N-1} U(\tfrac{k}{N}, x_k, \tfrac{x_{k+1}-x_k}{dt})\, dt \text{ s.t. } x_0 = b_0,\ x_N = b_1, \text{ or as} \tag{87}$$

$$\max_{a_k:k=0,\ldots,N-1} \sum_{k=0}^{N} U(\tfrac{k}{N}, \textstyle\sum_{j<k} a_j\, ds, a_k)\, dt \text{ s.t. } x_N = b_0 + \sum_{j<N} a_j\, ds = b_1. \tag{88}$$

The first formulation focuses our attention on the path that the state takes, $t \mapsto x_t$ for $t \in T$, the second focuses our attention on the pathh that the actions take. In looking at the associated first order conditions, each x_k appears twice, while each a_k appears in each of the summation terms from $k+1$ onwards.

Note that we are omitting the point 1 from the summation, but that this does not matter, at time 1, decisions have no effect because there is no future left for any action to change.

In a similar fashion, the second class of problems can be formulated as

$$\max_{a_k:k=0,\ldots,N-1} \sum_{k=0}^{N-1} U(\tfrac{k}{N}, x_k, a_k)\, dt \text{ s.t. } x_0 = b_0,\ x_N = b_1,\ a_k \in A$$
$$\text{and } x_{k+1} = x_k + f(\tfrac{k}{N}, x_k, a_k)\, dt,\ k = 0, \ldots, N-1. \tag{89}$$

Again, taking $f(t, x, a) \equiv a$ and $A = \mathbb{R}$ recovers the first class of problems.

9.6.2. *Necessary Conditions.* There are many ways to arrive at the necessary conditions for problem 87. The easiest is the most direct, at an optimal path x_k^*, $k = 1, \ldots, N-1$, the derivative of the objective function must equal 0. Since x_k appears three times in the summation, this yields, after removing the common factor of dt,

$$\frac{\partial U(\frac{k-1}{N}, x_{k-1}^*, \frac{x_k^*-x_{k-1}^*}{dt})}{\partial \dot{x}} \frac{1}{dt} + \frac{\partial U(\frac{k}{N}, x_k^*, \frac{x_{k+1}^*-x_k^*}{dt})}{\partial x} - \frac{\partial U(\frac{k}{N}, x_k^*, \frac{x_{k+1}^*-x_k^*}{dt})}{\partial \dot{x}} \frac{1}{dt} = 0. \tag{90}$$

Rearranging and supressing some arguments, this is

$$\frac{\partial U}{\partial x} = \frac{d}{dt} \frac{\partial U}{\partial \dot{x}}, \tag{91}$$

and this has to hold at every point along the optimal path.

9.7. **Some Examples of Using the Euler-Lagrange Conditions.**

9.7.1. *A Squared Law of Resistance.*

$$\min_{a_k:k=0,\ldots,N-1} \sum a_k^2\, dt \text{ s.t.} x_N = 0 + \sum_{j<N} a_j\, ds = B. \tag{92}$$

The Lagrangean is $L(a, \lambda) = \sum a_k^2 + \lambda(B - \sum_{j<N} a_j$, and from the FOC, one sees that at every k, $2a_k = \lambda$, that is, the a_k are constant. The E-L necessary conditions are, at every point along the optimal path,

$$0 = \frac{d}{dt} \frac{\partial U}{\partial \dot{x}}, \tag{93}$$

that is, $\dot{x}^*(t) = a^*(t)$ cannot change with t.

9.8. **Control Problems Bang-Bang Solutions.** So far the solutions $t \mapsto a_t^*$ have either been near continuous, or they have been wildly discontinuous. Intermediate between these are solutions where the graph of a_t^* looks like the graph of a piecewise continuous function in $V(S)$. When the solution bounces between extreme points of the constraint set, we call it a **bang-bang** solution. The idea is that the optimal way to control the process is to slam the controls between extremes at optimally chosen points in time, and the slamming results in a banging sound (at least inside our heads).

9.9. **Savings/Investment Problems.**

For any infinite n set $N = n!$, let $T = \{\frac{k}{N} : k \in \{0, 1, \ldots, N^2\}\}$ with t_k denoting the k'th time point in T, i.e. $t_k = \frac{k}{N}$, and dt, as usual, denoting the time increment $\frac{1}{N}$. Let $\Omega = \{-1, +1\}^T$ so that $|\Omega| = 2^{N^2+1}$, $\mathcal{F}$ the class of internal subsets of Ω, and let P be the uniform distribution, $P(A) = \frac{|A|}{2^{N^2+1}}$. For each $\omega \in \Omega$ and $k \in \{0, 1, \ldots, N^2\}$, define $\omega_k = \text{proj}_{t_k}(\omega)$.

10.1. **Two Versions of the Time Lines.** We have two ways to define the non-standard process for Brownian motion, one that specifies the value at each $t \in T$, and one that specifies the value for each t in the *-infinite, internal interval $[0, N^2]$. The second is just linear interpolation of the first. The first is defined at each $t_k \in T$ by

$$\chi_d(t_k, \omega) = \frac{1}{\sqrt{N}} \sum_{j=1}^{k} \omega_j, \tag{94}$$

the second is defined at each $t \in [0, N^2]$ by

$$\chi_c(t, \omega) = \frac{1}{\sqrt{N}} \sum_{j=1}^{\lfloor Nt \rfloor} \omega_j + (Nt - \lfloor Nt \rfloor)\omega_{j+1} \tag{95}$$

where for $x \in {}^*\mathbb{R}_+$, $\lfloor x \rfloor$ is the largest integer less than or equal to x.

We will now show that either of the following defines a Brownian motion on $(\Omega, L^(\mathcal{F}), L(P))$, for standard $t \in [0, \infty)$ and $\omega \in \Omega$, either

$$b(t, \omega) = {}^\circ\chi_d(t+, \omega) \text{ or } \beta(t, \omega) = {}^\circ\chi_c(t, \omega) \tag{96}$$

where for any standard $t \in [0, \infty)$, $t+$ is the first $t \in T$ greater than or equal to t, $t+ := \min\{t_k \in T : t \le t_k\}$.

10.2. **Showing Brownian-ness.** An implication of the proofs is that for a set of ω having probability 1, the paths $t \mapsto b(t, \omega)$ and equal to the paths $t \mapsto \beta(t, \omega)$.

Lemma 10.2.1. *For standard* $0 \le s < t$, $b(t, \cdot) - b(s, \cdot) \sim N(0, (t - s))$.

Proof. From the theory of characteristic functions, a random variable, Y, has a $N(0, \sigma)$ distribution iff $E\, e^{irY} = e^{-\frac{1}{2}r^2\sigma}$. Now,

$$\int e^{ir(b(t,\omega)-b(s,\omega)}\, dL(P)(\omega) \simeq \textstyle\int {}^*e^{ir(\chi_d(t+,\omega)-\chi_d(s+,\omega)}\, d(P)(\omega) \tag{97}$$

$$= \int {}^*e^{ir\frac{1}{\sqrt{N}}\sum_{k:s+<t_k\le t+}\omega_k}\, dP(\omega) \tag{98}$$

$$= \Pi_{k:s+<t_k\le t+} \textstyle\int {}^*e^{ir\omega_k}\, dP(\omega) \tag{99}$$

$$= \Pi_{k:s+<t_k\le t+} {}^*\cos\left(\frac{r}{\sqrt{N}}\right) \tag{100}$$

where the switch to $\cos(\cdot)$ comes from $e^{ix} = \cos(x) + i\sin(x)$, and the symmetry of the distribution of the ω_k, which implies that $E\, \sin(\omega_k) = 0$ because $\sin(x) = -\sin(-x)$. Now, the Taylor expansion of $\cos(x)$ is $\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + o(x^5)$. This means that $\Pi_{k:s+<t_k\le t+} {}^*\cos(\frac{r}{\sqrt{N}}) \simeq e^{\frac{1}{2}r^2(t-s)}$ as required. □

Lemma 10.2.2. *For any $s < t$, $s, t \in T$, $E\left(\chi_d(t,\cdot) - \chi_d(t,\cdot\right)^2 = (t-s)^2$ and $E\left(\chi_d(t,\cdot) - \chi_d(t,\cdot\right)^4 = 3(t-s)^2 - 2(t-s)dt$.*

Proof. For the standard parts, these are just the second and fourth central moments of a $N(0, t-s)$. Direct calculations, using the independence of increments to cancel out the odd power terms, delivers these results. □

Now, if $\omega_1 = \omega_2 = \cdots = \omega_N = 1$, then $\chi_d(t_N, \omega) = N \cdot \frac{1}{\sqrt{N}} = \sqrt{N} \simeq \infty$, for this ω, the time path has no standard part. The next result is crucial, it says that the probability that this kind of oddity happening is 0.

Theorem 10.2.3. $L(P)(\{\omega : t \mapsto b(t,\omega)$ *is continuous*$\}) = 1$.

The observation behind the proof is that a standard function $f : [0,\infty) \to \mathbb{R}$ is continuous iff for all integers n and k, for large enough m, we do not have an $s \in [\frac{i}{m}, \frac{i+1}{m}] \subset [0,k]$ such that $(f(\frac{i}{m}) - f(s))^4 \geq \frac{1}{n}$. This is so because a continuous function on $[0,\infty)$ is uniformly continuous on each $[0,k]$.

Proof. For each standard $k \in \mathbb{N}$, define $T_k = [0,k] \cap T$. For any standard triple of integers, m, n, k, here is a set that we would like to have small probability,

$$\Omega_{m,n,k} = \bigcup_{i\in\mathbb{N}} \left\{\omega : \left(\exists s \in (T_k \cap \left[\tfrac{i}{m}, \tfrac{i+1}{m}\right]\right) \left[|\chi_d(\tfrac{1}{m},\omega) - \chi_d(s,\omega)|^4 \geq \tfrac{1}{n}\right]\right\}. \tag{101}$$

We show that $\lim_{m\uparrow\infty} P(\Omega_{m,n,k}) = 0$, the limit being taken along $\mathbb{N}$.

The lovely (or sneaky) part of the argument is the observation that

$$\begin{aligned} P(\{\omega :&|\chi_d(\frac{1}{m},\omega) - \chi_d(s,\omega)|^4 \geq \frac{1}{n}\}) \\ &\leq 2 \cdot P(\{\omega : |\chi_d(\tfrac{1}{m},\omega) - \chi_d(\tfrac{i+1}{m},\omega)|^4 \geq \tfrac{1}{n}\}). \end{aligned} \tag{102}$$

This comes from a "reflection" argument for random walks. Pick any ω such that $|\chi_d(\frac{1}{m},\omega) - \chi_d(\frac{i+1}{m},\omega)|^4 < \frac{1}{n}$ but for some $s \in T \cap [\frac{i}{m}, \frac{i+1}{m}]$, $|\chi_d(\frac{1}{m},\omega) - \chi_d(s,\omega)|^4 \geq \frac{1}{n}$. For this ω, let s_ω be the smallest time in $T \cap [\frac{i}{m}, \frac{i+1}{m}]$ at which $|\chi_d(\frac{1}{m},\omega) - \chi_d(s,\omega)|^4 \geq \frac{1}{n}$. Look at the path that comes from switching the sign of each ω_k for $t_k \geq s_\omega$. The new ω' has exactly the same probability, and because $|\chi_d(\frac{1}{m},\omega) - \chi_d(\frac{i+1}{m},\omega)|^4 < \frac{1}{n}$, we know that $|\chi_d(\frac{1}{m},\omega') - \chi_d(\frac{i+1}{m},\omega')|^4 > \frac{1}{n}$. Hence, if we 'double count' the number of ω's with $|\chi_d(\frac{1}{m},\omega) - \chi_d(\frac{i+1}{m},\omega)|^4 \geq \frac{1}{n}$, we must have counted all of the ω's with $|\chi_d(\frac{1}{m},\omega) - \chi_d(s,\omega)|^4 \geq \frac{1}{n}$ for some $s \in T \cap [\frac{i}{m}, \frac{i+1}{m}]$.

Now we use a counting argument, the moments above, and Chebyshev,

$$P(\Omega_{m,n,k}) \leq 2 \cdot \sum_{i=0}^{km-1} P(\{\omega : |\chi_d(\frac{1}{m},\omega) - \chi_d(\frac{i+1}{m},\omega)|^4 \geq \frac{1}{n}\}) \tag{103}$$

$$\leq 2 \cdot \sum_{i=0}^{km-1} nE\,|\chi_d(\frac{i}{m},\cdot) - \chi_d(\frac{i+1}{m},\omega)|^4 \tag{104}$$

$$\leq 6n \cdot \sum_{i=0}^{km-1} \frac{i}{m^2} \leq \frac{6kn}{m}. \tag{105}$$

As $(6kn)/m \to 0$ as $m \uparrow \infty$, $L(P)(\cup_{n,k} \cap_m \Omega_{m,n,k}) = 0$. □

10.3. **Derivatives and Bounded Variation.** For a standard function $f : [0, \infty) \to \mathbb{R}$ and $t \in [0, \infty)$, the upper and lower right derivatives at t are

$$D^+ f(t) := \lim_{\epsilon \downarrow 0} \sup_{s \in (t, t+\epsilon)} \tfrac{f(s)-f(t)}{s-t} \tag{106}$$

$$D_+ f(t) := \lim_{\epsilon \downarrow 0} \inf_{s \in (t, t+\epsilon)} \tfrac{f(s)-f(t)}{s-t}. \tag{107}$$

If these two are equal and finite, f has a right derivative at t. The following tells us that Brownian motion paths are never differentiable.

Lemma 10.3.1. *$L(P)(\{\omega : D^+ b(t, \omega) = \infty\}) = L(P)(\{\omega : D_+ b(t, \omega) = -\infty\}) = 1$.*

The variation of a standard function $f : [0, \infty) \to \mathbb{R}$ over an interval $[s, t]$ is the supremum of the sums of the form $\sum |f(t_i) - f(s_i)|$ where $s = s_1 < t_1 \le s_2 < t_2 \le \cdots \le s_I < t_I = t$. A function has locally bounded variation if its variation over any compact interval is finite, and it has some locally bounded variation if it has bounded variation of some compact non-degenerate interval The following tells us that Brownian motion paths have, with probability 1, infinite variation over every compact interval.

Lemma 10.3.2. *$L(P)(\{\omega : b(t, \omega)$ has some locally bounded variation$\}) = 0$.*

10.4. **Itô's Lemma.** The relation between bounded variation and integral is taught in high-end calculus classes. For a piecewise continuous function $h : [0, 1] \to \mathbb{R}$ and a $G : [0, 1] \to \mathbb{R}$ having bounded variation on $[0, 1]$, the Riemann-Stjeltjes integral is

$$\int_0^1 h(t)\, dG(t) := \lim \sum \left(\sup_{x \in [s_i, t_i]} h(x)\right) (G(t_i) - G(s_i)) \tag{108}$$

with the limits being taken over the same kinds of s_i, t_i pairs. The basic result is that $\int_0^1 h(t)\, dG(t) = \lim \sum \left(\inf_{x \in [s_i, t_i]} h(x)\right) (G(t_i) - G(s_i))$ for all piecewise continuous $h(\cdot)$ and does not depend on the sequence of subdivisions iff $G(\cdot)$ is of bounded variation.

Despite this, we would like to be able to define

$$\int_0^1 h(t)\, db(t, \omega) \text{ and } \int_0^1 h(t, \omega)\, db(t, \omega). \tag{109}$$

References

[Bergin and Macleod, 1993] Bergin, J. and Macleod, W. B. (1993). Continuous time repeated games. *International Economic Review*, 34(1):21–37.

[Blume et al., 1991a] Blume, L., Brandenburger, A., and Dekel, E. (1991a). Lexicographic probabilities and choice under uncertainty. *Econometrica*, 59(1):61–79.

[Blume et al., 1991b] Blume, L., Brandenburger, A., and Dekel, E. (1991b). Lexicographic probabilities and equilibrium refinements. *Econometrica*, 59(1):81–98.

[Clarke, 2013] Clarke, F. (2013). *Functional analysis, calculus of variations and optimal control*, volume 264 of *Graduate Texts in Mathematics*. Springer, London.

[Corbae et al., 2009] Corbae, D., Stinchcombe, M. B., and Zeman, J. (2009). *An introduction to mathematical analysis for economic theory and econometrics*. Princeton University Press, Princeton, NJ.

[Fleming and Rishel, 1975] Fleming, W. H. and Rishel, R. W. (1975). *Deterministic and stochastic optimal control*. Springer-Verlag, Berlin. Applications of Mathematics, No. 1.

[Fudenberg and Levine, 1986] Fudenberg, D. and Levine, D. (1986). Limit games and limit equilibria. *Journal of Economic Theory*, 38(2):261–279.

[Fudenberg and Levine, 1993] Fudenberg, D. and Levine, D. K. (1993). Self-confirming equilibrium. *Econometrica*, 61(3):523–545.

[Fudenberg and Tirole, 1985] Fudenberg, D. and Tirole, J. (1985). Preemption and rent equalization in the adoption of new technology. *The Review of Economic Studies*, 52(3):383–401.

[Fudenberg and Tirole, 1991] Fudenberg, D. and Tirole, J. (1991). *Game theory*. MIT Press, Cambridge, MA.

[Kakutani, 1948] Kakutani, S. (1948). On equivalence of infinite product measures. *Ann. of Math. (2)*, 49:214–224.

[Liberzon, 2012] Liberzon, D. (2012). *Calculus of variations and optimal control theory*. Princeton University Press, Princeton, NJ. A concise introduction.

[Lindstrøm, 1988] 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.

[Manelli, 1996] Manelli, A. M. (1996). Cheap talk and sequential equilibria in signaling games. *Econometrica*, 64(4):917–942.

[McShane and Warfield, 1969] McShane, E. J. and Warfield, Jr., R. B. (1969). Addenda and corrigenda to "On Filippov's implicit functions lemma". *Proc. Amer. Math. Soc.*, 21:496–498.

[Nelson, 1987] Nelson, E. (1987). *Radically elementary probability theory*, volume 117 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ.

[Perry and Reny, 1993] Perry, M. and Reny, P. J. (1993). A non-cooperative bargaining model with strategically timed offers. *Journal of Economic Theory*, 59(1):50–77.

[Simon and Stinchcombe, 1989] Simon, L. K. and Stinchcombe, M. B. (1989). Extensive form games in continuous time: pure strategies. *Econometrica*, 57(5):1171–1214.

[Simon and Stinchcombe, 1995] Simon, L. K. and Stinchcombe, M. B. (1995). Equilibrium refinement for infinite normal-form games. *Econometrica*, 63(6):1421–1443.

[Simon and Zame, 1990] Simon, L. K. and Zame, W. R. (1990). Discontinuous games and endogenous sharing rules. *Econometrica*, 58(4):861–872.

[Stinchcombe, 1992] Stinchcombe, M. B. (1992). Maximal strategy sets for continuous-time game theory. *J. Econom. Theory*, 56(2):235–265.

[Stinchcombe, 2005] Stinchcombe, M. B. (2005). Nash equilibrium and generalized integration for infinite normal form games. *Games Econom. Behav.*, 50(2):332–365.