Notes on Measure Theory and Markov Processes

Diego Daruich*

May 2, 2014

1 Preliminaries

1.1 Motivation

The objective of these notes will be to develop tools from measure theory and probability to allow us to analyze the dynamics of economies with uninsurable income risk. In these models, agents are heterogeneous in the vector of individual states, and we need some way to describe this heterogeneity. The natural way is to use two objects:

1. A *probability measure* to keep track of the measure of agents in various regions of the state space.
2. A *transition function* to describe the law of motion of this measure.

The definitions and results presented will follow closely from Stokey and Lucas (1989, hereafter "SL") and Ljungqvist and Sargent (2004, hereafter "LS").

1.2 Definitions From Measure Theory

Before dealing with the results, it will be useful to characterize a number of important objects from measure theory.

Definition. A **σ-algebra** $\mathcal{F}$ on a set S is defined to be a collection of sets such that

1. $\emptyset \in \mathcal{F}$.
2. If $A \in \mathcal{F}$, then $S \backslash A \in \mathcal{F}$ (i.e., $\mathcal{F}$ is closed under complements).
3. If $A_i \in \mathcal{F}$ for each i, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ (i.e., $\mathcal{F}$ is closed under countable unions).

*E-mail: diedar@gmail.com. Based on Dan Greenwald's notes, Stokey and Lucas (1989) and Ljungqvist and Sargent (2013).

This is the collection of subsets where measures will be defined.

Definition. A **measurable space** is defined to be a pair $(S, \mathcal{F})$, where S is a set, and $\mathcal{F}$ is a σ-algebra.

Definition. Let $(S, \mathcal{F})$ be a measurable space. A **measure** on $(S, \mathcal{F})$ is a function $\mu : \mathcal{F} \to \bar{R}_+$ (the non-negative portion of the extended real line), such that:

1. $\mu(\emptyset) = 0$.
2. If $\{A_i\}$ is a countable disjoint sequence of sets in $\mathcal{F}$, then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

   This property is known as countable additivity.

If, in addition, $\mu(S) < \infty$, μ is said to be a **finite measure**. If $\mu(S) = 1$, then μ is said to be a **probability measure**.

Definition. A **measure space** is a triple $(S, \mathcal{F}, \mu)$, where $(S, \mathcal{F})$ is a measurable space, and μ is a measure on $(S, \mathcal{F})$. If μ is a probability measure, then we say that $(S, \mathcal{F}, \mu)$ is a **probability space**.

Definition. Given a measurable space $(S, \mathcal{F})$, a function $f : S \to \mathbb{R}$ is called **$\mathcal{F}$-measurable** if

$$\{s \in S : f(s) \leq a\} \in \mathcal{F}$$

for all $a \in \mathbb{R}$.

These definitions are mostly intended to clarify the statements and conditions of results in further sections. In general, the basic results of measure theory will not be covered in these notes, but the interested reader should refer to SL, Chapter 7.

1.3 Lebesgue Integration

Given a measure μ, and an "integrable" function g, measure theory contains a notion of a "Lebesgue" integral in the form

$$\int_A g(s)\, \mu(ds).$$

In shorthand, this integral is sometimes written as

$$\int_A g\, d\mu.$$

A full treatment of the definition and properties of this integral goes far beyond the scope of these notes (see SL, Chapter 7 for more details). However, when the integral is taken with respect to a probability measure (e.g., an expectation), and g satisfies some regularity conditions, some intuition can be gained by comparing this Lebesgue integral to the more familiar Stieltjes and Riemann integrals.[1] Assume that $A = [a, b]$, that g is "well-behaved," and that p is a probability measure that admits a distribution function F and a density f. Then we have

$$\int_A g(s)\, p(ds) = \int_a^b g(s)\, dF(s) = \int_a^b g(s) f(s)\, ds.$$

For a discrete probability distribution with mass function f, and $A = \{s_i\}$ we would similarly obtain

$$\int_A g(s)\, p(ds) = \sum_{i=1}^{\infty} g(s_i) f(s_i).$$

Of course, not all probability measures and functions will have this convenient form. My hope is that this relationship should hopefully take some of the mystery out of the Lebesgue integral notation.[2] What is the benefit of the Lebesgue integral? The Riemann integral is problematic when requiring interchanging limit processes and integral signs (if interested notice that the Monotone convergence theorem fails).

2 Markov Processes

2.1 The Transition Function

For our analysis, we will restrict our attention to Markov processes, in which the probability of events in any period depend only on the state in the previous period.[3] In thise case, probabilities can be completely described by a transition function.

Definition. A **transition function** on a measure space $(S, \mathcal{F})$ is a function $Q : S \times \mathcal{F} \to [0, 1]$ such that

1. For each $s \in S$, $Q(s, \cdot)$ is a probability measure on $(S, \mathcal{F})$.
2. For each $A \in \mathcal{F}$, $Q(\cdot, A)$ is a $\mathcal{F}$-measurable function.

$Q(s, A)$ is a conditional probability function, i.e. the probability of the state moving from s in one period to an element of A in the next period. For a Markov process, the dynamics are completely described by the transition function. The extreme convenience of Markov processes therefore follows not only from avoiding the need to condition on states more than one period in the past, but more importantly from the fact that the same transition function can be used in all periods, making each period's problem symmetric.

This is in stark contrast with the general case, in which we would have to use a new probability measure $p_t(\cdot|s^{t-1})$ in each period, making the problems asymmetric across periods.

[1] Recall that the Stieltjes and Riemann integral is computed by taking successively finer grids on the "x-axis" for the step function that is greater/less than the objective function. If both of them coincide, the objective function is said to be Riemann integrable. On the other hand, the Lebesgue integral is computed by taking finer grids on the "y-axis" and (possibly) "weighting" by a measure.

[2] For a case when the two differ, look at the Dirichlet function, in which case the Riemann integral does not exist.

[3] For a more rigorous discussion of Markov processes, and stochastic processes in general, see SL, Chapter 8.

2.2 Economic Example

In a typical macro model, Q is determined by exogenous laws of motion for the states, and by policy functions. For a model with assets and labor income, we might have $S = A \times Y$, and so Q would depend on $a'(a, y)$ and $\pi(y', y)$.

For a simple example, suppose that labor income can take only two values $\{y_L, y_H\}$, that that wealth can take only two values, $\{a_L, a_H\}$. Assume a Markov chain on labor income $\pi(y', y)$ and a decision rule

$$a'(a_i, y_j) = a_j$$

for $i, j = L, H$. In this case, Q would be given by

$$Q\Big((a_{i_t}, y_{j_t}), (a_{i_{t+1}}, y_{j_{t+1}})\Big) = \pi(y_{j_{t+1}}, y_{j_t}) \cdot \mathbf{1}_{\{a'(a_{i_t}, y_{j_t}) = a_{i_{t+1}}\}}.$$

2.3 Markov Operators

Given a Markov process on a measurable space $(S, \mathcal{F})$, and the associated transition function Q, we can define two important operators, known as **Markov operators**. Define $M_+(S, \mathcal{F})$ to be the space of non-negative measurable functions on $(S, \mathcal{F})$. Then we can define the Markov operator $T : M_+(S, \mathcal{F}) \to M_+(S, \mathcal{F})$ by

$$Tf(s) = \int_S f(s')Q\left(s, ds'\right).$$

$Tf(s)$ represents the conditional expectation of f (assumed to be measurable) next period, given that the current state is s. We could also have defined $T : B(S, \mathcal{F}) \to B(S, \mathcal{F})$, where $B(S, \mathcal{F})$ is the space of bounded measurable functions on $(S, \mathcal{F})$.

Similarly, let $\Lambda(S, \mathcal{F})$ be the set of probability measures on $(S, \mathcal{F})$. Then we can define the other Markov operator $T^* : \Lambda(S, \mathcal{F}) \to \Lambda(S, \mathcal{F})$ by

$$T^*\lambda(A) = \int_S Q(s, A)\,\lambda(ds).$$

$T^*\lambda$ is the probability measure next period, given that current values of the state are drawn according to the probability measure λ.

Proofs that T and T^* are in fact the self-maps that they are stated to be can be found in SL, Chapter 8.

In our models we will often use λ to represent the proportion of agents in each state, rather than the probability that a single agent is in that state. By a "law of large numbers" argument, these definitions should be equivalent over a continuum of agents.

2.4 Application to Dynamic Programming

Before moving on, I would like to mention one application of this Markov process framework to the standard dynamic programming problem. Generally, in class we treat stochastic dynamic programming problems exactly like deterministic dynamic programming problems, and apply the same theorems and results. While this is usually fine, it is worth noting that we can easily deal with the stochastic problem under a Markov process in full rigor, if we apply the results from SL, Chapter 9. The following theorem is particularly useful.

Theorem (SL 9.6)**.** Let $(X, \mathcal{X})$ and $(Z, \mathcal{Z})$ be measurable spaces, let Q be a transition function on $(Z, \mathcal{Z})$, let Γ be a correspondence on $X \times Z$, let $F : A \to \mathbb{R}$, for some set A, and let $\beta \in \mathbb{R}$. Further, assume that

1. X is a convex Borel set in $\mathbb{R}^l$ and $\mathcal{X}$ is the Borel σ-algebra on X.[4]
2. One of the following conditions holds:
   a. Z is a countable set and $\mathcal{Z}$ is the σ-algebra containing all subsets of Z.
   b. Z is a compact Borel set in $\mathbb{R}^k$, $\mathcal{Z}$ is the Borel σ-algebra on Z, and Q has the Feller property.[5]
3. Γ is nonempty, compact-valued, and continuous.
4. F is bounded and continuous, and $\beta \in (0, 1)$.

Let $CB(S)$ be the space of bounded continuous functions on $S = X \times Z$, define the operator T on $CB(S)$ by[6]

$$(Tf)(x, z) = \sup_{y \in \Gamma(x,z)} \left\{ F(x, y, z) + \beta \int f(y, z')Q(z, dz') \right\}.$$

Then

1. $T : CB(S) \to CB(S)$, where CB is the space of bounded continuous functions on $S = X \times Z$.
2. T has a unique fixed point $v \in CB(S)$, and for any $v_0 \in CB(S)$,

$$||T^n v_0 - v|| \leq \beta^n ||v_0 - v||$$

   for all $n \in \mathbb{N}$.
3. The correspondence $G : S \to X$ defined by

$$G(x, z) = \left\{ y \in \Gamma(x, z) : v(x, z) = F(x, y, z) + \beta \int v(y, z')Q(z, dz') \right\}$$

   is nonempty, compact-valued, and upper hemicontinuous.

For more results on stochastic dynamic programming, see SL, Chapter 9.

[4]A Borel set is any set in a topological space that can be formed from open sets (or, equivalently, from closed sets) through the operations of countable union, countable intersection, and relative complement. For a topological space X, the collection of all Borel sets on X forms a σ-algebra, known as the Borel algebra or Borel -algebra. The Borel algebra on X is the smallest σ-algebra containing all open sets (or, equivalently, all closed sets). Borel sets are important in measure theory, since any measure defined on the open sets of a space, or on the closed sets of a space, must also be defined on all Borel sets of that space.

[5]The Feller property is defined in Section 3.2.

[6]This T operator should not be confused with the Markov operator T of Section 2.3.

3 Invariance and Convergence

3.1 Invariant Distributions

An important concept for our analysis will be that of an invariant distribution, also known as a stationary distribution.

Definition. Given a Markov process on $(S, \mathcal{F})$ with associated transition function Q, a probability measure λ is called **invariant**, if it is a fixed point of T^*, i.e. if

$$\lambda(A) = \int_S Q(s, A)\, \lambda(ds)$$

for all $A \in \mathcal{F}$.

This property of an invariant distribution means that if an economy governed by a Markov process begins where the probability distribution of states across agents is invariant, then it will maintain that same in all future periods.

Another reason why invariant distributions are important is that if the invariant distribution is unique, then probabilities and moments with respect to that distribution can be considered unconditional.

In general, for a given model we will be concerned with three questions with regard to invariant distributions.

1. Does an invariant distribution exist?
2. If so, is it unique?
3. How can an invariant distribution be calculated?

The remainder of this section will be concerned with the answers to these questions. Material will be mainly drawn from SL, Chapters 11 and 12, to which the interested reader should refer for more detail.

3.2 Weak Convergence Approach

One approach to answering the questions above uses the notion of "weak convergence" of probability distributions.

Definition. A sequence of $\{\lambda_n\}$ of probability distributions on a measurable space $(S, \mathcal{F})$ is said to **converge weakly** to a probability measure λ (denoted $\lambda_n \Rightarrow \lambda$) if

$$\lim_{n\to\infty} \int f\, d\lambda_n = \int f\, d\lambda$$

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

While this notion of convergence will not appear directly in all of the results to follow, most of the proofs are based on it. For a full treatment, see SL, Chapter 12.

To establish existence and uniqueness of an invariant distribution under this approach, we will need the following definitions.

Definition. A transition function Q on $(S, \mathcal{F})$ satisfies the **Feller property** if the associated Markov operator T maps $CB(S, \mathcal{F})$ onto itself, where $CB(S, \mathcal{F})$ is the space of continuous bounded measurable functions on $(S, \mathcal{F})$.

Hence, this puts continuity constraints on the transition function. All $\mathbb{R}^n$-valued processes with stationary independent increments (i.e., Levy processes) are Feller. However, notice that in Economics the transition function is usually an endogenous object (it depends on the policy functions), which can complicate proving the presence of this property.

Definition. A transition function Q on $(S, \mathcal{F})$ is **monotone** if the associated Markov operator T maps the space of increasing measurable functions onto itself.

Definition. Assume $S = [a, b] \subseteq \mathbb{R}^n$.[7] A transition function Q on $(S, \mathcal{F})$ satisfies the **monotone mixing condition** if there exist $\hat{s} \in S$, $\varepsilon > 0$, and $N \geq 1$ such that $Q^N(a, [\hat{s}, b]) \geq \varepsilon$ and $Q^N(b, [a, \hat{s}]) \geq \varepsilon$.

With these definitions in hand, we can obtain the following powerful results.

Theorem (SL 12.10)**.** If $S \subseteq \mathbb{R}^l$ is compact, and Q has the Feller property, then there exists a probability measure that is invariant under Q.

Theorem (SL 12.12)**.** Let $S = [a, b] \subseteq R^l$. If Q is monotone, has the Feller property, and satisfies the monotone mixing condition, then P has a unique invariant probability measure, and $(T^*)^n \lambda_0 \Rightarrow \lambda^*$ for all $\lambda_0 \in \Lambda(S, \mathcal{F})$.

This allows us to determine existence and uniqueness of an invariant distribution, and also a method to calculate it, through repeated application of the T^* operator.

3.3 Strong Convergence Approach

Another approach, found in SL, Chapter 11, is based on the notion of "strong convergence." This approach will yield similar results to Section 3.2 under slightly different properties, and is rather technical. Therefore, the reader may want to skip ahead to 4 to explore the highly useful special case of Markov Chains.

A sequence of probability measures $\{\lambda_n\}$ converges strongly to a probability measure λ if $\lambda_n \Rightarrow \lambda$, and in addition, the rate of convergence of $\int f \, d\lambda_n$ to $\int f \, d\lambda$ is uniform for all $f \in B(S, \mathcal{F})$ such that $||f|| \leq 1$ under the usual sup norm.

As in the previous section, this definition is not particularly important for understanding the results, although this notion of convergence is used to obtain them. For the interested reader, strong convergence has a more intuitive expression with regard to the "total variation norm" (see SL, Chapter 11 for details).

[7] An "interval" $[a, b]$ on $\mathbb{R}^n$, with $a, b \in \mathbb{R}^n$ is the rectangle $[a_1, b_1] \times \cdots \times [a_n, b_n]$.

Under this approach, we need to add an additional definition.

Definition. A transition function satisfies **Condition D** if there exists a finite measure ϕ on $(S, \mathcal{F})$, an integer $N \geq 1$, and a number $\varepsilon > 0$, such that if $\phi(A) \leq \varepsilon$, then $Q^N(s, A) \leq 1 - \varepsilon$ for all $s \in S$.

We can now state the following results.

Theorem (SL 11.9)**.** Let $(S, \mathcal{F})$ be a measurable space, and let Q be a transition function on $(S, \mathcal{F})$ that satisfies Condition D. Then

1. S can be partitioned into a transient set and M ergodic sets, where $1 \leq M \leq \phi(S)/\varepsilon$.
2. The limit
$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} (T^*)^n \lambda_0$$
exists for each $\lambda_0 \in \Lambda(S, \mathcal{F})$, and for each λ_0, this limit is an invariant distribution.
3. There are M invariant measures corresponding to the M ergodic sets, and every invariant measure of T^* can be written as a convex combination of these.

The concepts of a transient and ergodic set have not been defined in these notes, but the important point to take away from the first and third results is that this theorem implies a finite-dimensional space of invariant distributions.

Theorem (SL 11.10)**.** Let $(S, \mathcal{F})$ be a measurable space, let Q be a transition function on $(S, \mathcal{F})$, and assume that Q satisfies Condition D for finite measure ϕ. Suppose in addition that if $\phi(A) > 0$, then for each $s \in S$ there exists $n \geq 1$ such that $Q^n(s, A) > 0$. Then

1. S has only one ergodic set.
2. T^* has only one invariant measure, λ^*.
3. For each $\lambda_0 \in \Lambda(S, \mathcal{F})$, we have
$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} (T^*)^n \lambda_0 = \lambda^*.$$

As before, the reader should either ignore the first result above, or refer to SL, Chapter 11. This theorem yields uniqueness of the invariant distribution and a method for its calculation, under different conditions than those used in Section 3.2.

4 Special Case: Markov Chains

4.1 Markov Chains

A particularly useful and simple type of Markov process is a Markov chain, in which the state space S is a finite set $\{s_1, \ldots, s_n\}$. In this case, the transition function can be described by a transition matrix P where

$$P_{ij} = Q(s_i, s_j) = \Pr(s' = s_j | s = s_i).$$

It is clear that $P_{ij} \geq 0$ for all i, j and that we must have $\sum_{j=1}^{n} P_{ij} = 1$ for all i. Many powerful results can be obtained under this assumption, which will have parallels in both Sections 3.2 and 3.3

4.2 Invariant Distributions Under Markov Chains

A probability distribution on a Markov chain is defined by $\pi = (\pi_1, \ldots, \pi_n)'$. In this case, the Markov operator T^* is defined by

$$T^*\pi = P'\pi$$

and so a distribution is invariant if and only if $\pi = P'\pi$. Therefore, a distribution is invariant if and only if it is an eigenvector of P' associated with a unit eigenvalue.

This property makes invariant distributions extremely easy to calculate numerically. In particular, we can do an eigenvalue decomposition of P' (for example, using Matlab's function `eig`) to calculate the eigenvalues of P', and the associated eigenvectors. While the eigenvectors are not unique, they span the space of eigenvectors of P'. Therefore, the space of invariant distributions is spanned by the space of eigenvectors associated with unit eigenvalues of P', once they have been scaled so that they are positive and sum to 1. This makes it extremely easy to find a basis for the space of invariant distributions, In addition, this method makes it extremely easy to determine if a unique invariant distribution exists, a sufficient condition is that this decomposition yields exactly one eigenvector associated with the unit eigenvalue.

A nice result can be stated as follows.

Theorem (SL 11.1)**.** Let P be the transition matrix for a Markov chain on the finite state space S. Then the sequence

$$\frac{1}{N}\sum_{n=0}^{N-1} P^n$$

converges to a stochastic matrix Q. Each row of Q is an invariant distribution, so $\pi_0 Q$ is an invariant distribution for each $\pi_0 \in \Lambda^l$, and every invariant distribution of P is a convex combination of the rows of Q.

This is generally not as good of a method for calculating the invariant distributions as the one given earlier, but does provide a proof of existence.

An additional powerful result regarding invariant distributions requires the following definition.

Definition. Let π_∞ be the unique invariant distribution for a Markov chain with transition matrix P. If for all distributions π_0 it is true that $(P')^t\pi_0$ converges to π_∞, then we say that the Markov chain is **asymptotically stationary** with a unique invariant distribution.

The result can now be stated.

Theorem (LS 2.2.2 or SL 11.2)**.** Let P be a stochastic matrix for which $P_{ij}^n > 0$ for all i, j under some value of n (note, this should be the *same* value for all i, j). Then the Markov chain defined by P has a unique invariant distribution π^*, and the process is asymptotically stationary.

Using the previous notation, now each row of Q is equal to π^*, and hence $\pi_0 Q = \pi^*$.[8] A Markov chain where the initial distribution is invariant is called a **stationary** Markov chain. For notation, we express a stationary Markov chain as the pair (P, π).

4.3 Invariant Functions and Ergodicity

Consider a function that depends on the state of a Markov chain (i.e. a random variable). Since the state space is finite, we can represent each random variable $y_t(s)$ as a vector $\bar{y}$, so that $\bar{y}_i = f(s_i)$. This convenient notation allows us to write

$$\mathbb{E}[y_{t+k}|s_t] = P^k\bar{y}$$

where the left-hand side is a vector in which the ith element is given by $\mathbb{E}[y_{t+k}|s_t = s_i]$.

Some nice results can be obtained using the notion of invariant functions or random variables, which is defined below.

Definition. A random variable $y_t = \bar{y}s_t$ is said to be **invariant** if $y_t = y_0$ for all $t \geq 0$ and all realizations of s_t.

Note that this means that an invariant random variable is an eigenvalue of P associated with the unit eigenvalue, and can be calculated in a manner identical to the invariant distributions in Section 4.2, by replacing P' with P.

Note also than an invariant function can still depend on the state, because we have not set any restrictions on y_0. In particular, if $P = I_n$ (so that no matter what state you start in, you stay there forever), *every* function is invariant. However, Markov chains like this tend to be somewhat "pathological," and if we rule out these cases we can obtain a final powerful result.

Definition. Let (P, π) be a stationary Markov chain.[9] The chain is said to be **ergodic** if the only invariant functions are constant with probability 1, i.e. $\bar{y}_i = \bar{y}_j$ for all i, j such that $\pi_i > 0$ and $\pi_j > 0$.

Theorem (LS 2.2.5)**.** Let $\bar{y}$ define a random variable on a stationary and ergodic Markov chain (P, π). Then

$$\frac{1}{T}\sum_{t=1}^{T} y_t \to \mathbb{E}[y_0]$$

with probability 1, where the expectation is taken with respect to the invariant distribution π.

[8]For a result on uniform convergence with Markov Chains see SL 11.4 Theorem.

[9]In other words, $\pi' = \pi' P$.

References

[1] Ljundqvist, L. and Sargent, T.J., *Recursive Macroeconomic Theory*. MIT Press, 2nd Edition, 2004.

[2] Stokey, N.L. and Lucas, R.E. and Prescott, E.C., *Recursive Methods in Economic Dynamics*. Harvard University Press, 1989.