Hansen Lecture Notes: Stochastic Processes, Stationarity and Ergodicity, and the Birkhoff LLN

John Heilbron

Winter 2017

1 Introduction

- The Law of Large Numbers (LLN) is the fundamental statistical result stating that a sample average converges to a population mean. It is used to generate consistent estimators of unknown model parameters from econometric data.
- The standard LLN applies to independent and identically distributed (iid) observations. But we believe that elements of a time-series are generated in a dependent fashion, that today's events shape tomorrow's. And the standard LLN does not hold under this alternative assumption.
- Stochastic process theory is a framework for describing how observations are generated that accounts for the dependence we believe to exist in time-series data. Careful study yields conditions under which we may revive a variant of the LLN.
- The required conditions are stationarity and ergodicity. Though these are abstract and unfamiliar-sounding, some intuition comes from the fact that they partially resurrect properties of sampling data obtained from iid distributions:
  - Iid sampling distributions are identical ex-ante, before any observation has been realized, and in-media, after some but not all observations have been realized. Stationarity implies sampling distributions that are identical ex-ante, before any observation has been realized.
  - Iid sampling distributions are the sampling distributions of each next observation conditioned on prior realizations and therefore also the sampling distribution, on average, of the next observation given prior realizations. Ergodicity ensures that ex-ante identical sampling distributions are the sampling distribution, on average, over infinite time, of the next observation given prior realizations.

2 Reading the Math

- Caveat: This section is non-technical and may contain errors. It is only intended to convey some intuition that helped me make good enough sense of unfamiliar notation.

- (Def.) A set, Ω, is a collection of elements, ω.

  - (Def.) A subset, $\Lambda \subseteq \Omega$, is a set comprised of points in the original set, $\Lambda = \{\omega \mid \omega \in \Omega, ...\}$.
  - (Def.) Given a set, Ω, the power set, 2^{Ω}, is the set of all subsets of the original set, $2^{\Omega} = \{\Lambda \mid \Lambda \subseteq \Omega\}$.
  - (Def.) Given a set, Ω, a sigma algebra, $\mathcal{F}$, is a set comprised of subsets of the original set that are closed under intersection and unions (and some other details), $\mathcal{F} = \{\Lambda \mid \Lambda \subseteq \Omega, \Lambda_1 \cup \Lambda_2 \in \mathcal{F}, \Lambda_1 \cap \Lambda_2 \in \mathcal{F}\}$.
  - (Def.) Given a set, Ω, the Borel sigma algebra, , is a sigma algebra that contains "almost all" subsets of the original set. (N.B. The Borel sigma algebra sort of approximates the power set, which, for technical reasons, may or may not itself be a sigma algebra.)
  - (Def.) Given a set, Ω, a filtration, $\{\mathcal{F}_t\}_t$, is a sequence of sigma algebras of the original set with the property that $\mathcal{F}_t \subseteq \mathcal{F}_{t'}\ \forall t' \geq t$

- A probability measure on the set X, $Q : \mathcal{B} \to [0,1]$, is a function that takes most subsets of X (i.e. elements of the sigma algebra, $\mathcal{B}$) and returns a size between 0 and 1.

  - $Q(X) = 1$, $Q(\emptyset) = 0$, and $Q(b_1 \cup b_2) = Q(b_1) + Q(b_2) - Q(b_1 \cap b_2)$.
  - When $X = \mathbb{R}^n$ (and $x \notin \mathcal{B}$ is not an input of Q), the quantity $Q(dx)$ denotes the infinitesimally small but non-zero measure of the infinitesimally small subset of X located at x.

- A random variable (X, $\mathbb{R}^n$, Pr) is a variable, X, on a set, $\mathbb{R}^n$, with an associated probability measure, Pr, but we often write X and suppress the other elements.

  - Given r.v.s X and Y, $E[Y \mid X]$ is also a r.v., the value of which depends on the value of X.
  - Given r.v. Y and the sigma algebra of set X, $\mathcal{B}$, $E[Y \mid \mathcal{B}]$ is a r.v., the value of which depends on the realization of a subset of X, $b \in \mathcal{B}$.
  - For a random variable, a PDF, a CDF, and a probability measure encode the same information: $Q(dx) = f(x)dx$; $Q(b) = \int_{x \in b} f(x)dx$; $f(x) = F'(x)$; $F(x) = \int_{s=-\infty}^{x} f(s)ds$

- An operator, $\mathbb{T} : \{f \mid .\} \to \{f \mid .\}$, is a mapping that takes a function and returns a function. Its domain and range are sets of functions.

- The expression for conditional expectation, $E[f(Y) \mid X]$, is an operator where the input function is $f(.)$ and the output function is $g(x) = E[f(Y) \mid X = x]$.
- An operator, $\mathbb{T}$, may be used to define a functional equation, e.g. $\mathbb{T}f = f$, the solution to which, if it exists, is a function.

3 Stochastic Processes, Markov Processes, Discrete State Markov Chains

- (Def.) A stochastic process, $(\{X_t\}_{t=0}^{\infty}, S, X, Pr)$, is a sequence of random variables in observation space, $\{X_t\}_{t=0}^{\infty} \in \mathbb{R}^n$, and an implied sequence of associated probability measures, $\{Pr(X_t \in .)\}_{t=0}^{\infty}$, both derived from a transformation, $S : \Omega \to \Omega$, an observation function, $X : \Omega \to \mathbb{R}^n$, and measure, $Pr : \mathcal{F} \to \mathbb{R}$.
  - For some unobserved state, ω, the transformation, S, describes a sequence of unobserved states, $\{S^t(\omega)\}_{t=0}^{\infty}$.
  - The observation function, X, maps this sequence of unobserved states, $\{S^t(\omega)\}_{t=0}^{\infty}$, into a sequence of observations, $\{X_t\}_{t=0}^{\infty}$, by $X_t = X_t(\omega) = X(S^t(\omega))$.
  - The measure on event space, $Pr : \Omega \to \mathbb{R}$, induces the probability measure on each observation X_t, $Pr(X_t \in b) = Pr\{\omega \mid X_t(\omega) \in b\}$.
- (Def.) A Markov process, $(\{X_t\}_{t=0}^{\infty}, Q_0, P)$, is a sequence of random variables in observation space, $\{X_t\}_{t=0}^{\infty} \in \mathbb{R}^n$, and an implied sequence of associated probability measures, $\{Pr(X_t \in .)\}_{t=0}^{\infty}$, the latter derived from an initial measure, $Q_0 : \mathcal{B} \to [0,1]$, and a conditional measure, $P : \mathcal{B} \times \mathbb{R}^n \to [0,1]$.
  - The initial measure, Q_0, and conditional measure, P, describe a sequence of measures, $\{Q_t\}_{t=0}^{\infty}$, by $Q_{t+1}(b) = \int P(b \mid x) Q_t(dx) \; \forall b \in \mathcal{B}$.
  - The probability measure on X_t is described by the corresponding measure in the sequence $\{Q_t\}_{t=0}^{\infty}$, $Pr(X_t \in b) = Q_t(b) \; \forall b \in \mathcal{B}$.
- (Def.) A discrete state Markov chain, $(\{X_t\}_{t=0}^{\infty}, q_0, \mathbb{P})$, is a sequence of discrete random variables in the standard basis, $\{X_t\}_{t=0}^{\infty} \in \{e_i\}_{i=1}^{n}$ where $e_i' = [0...1...0]$, and an implied sequence of associated probability mass functions, $\{\{Pr(X_t = e_i)\}_{i=1}^{n}\}_{t=0}^{\infty}$, the latter derived from an initial distribution, q_0 s.t. $q_0'\underline{1} = 1$ and $q_0' e_i \geq 0 \; \forall i$, and a transition matrix, $\mathbb{P}_{n \times n}$ s.t. $\mathbb{P}\underline{1} = \underline{1}$ and $e_i' \mathbb{P} e_{i'} = \mathbb{P}_{(i,i')} \geq 0 \; \forall i, i'$.
  - The initial distribution, q_0, and transition matrix, $\mathbb{P}$, describe a sequence of distributions, $\{q_t\}_{t=0}^{\infty}$, by $q_{t+1}' = q_t' \mathbb{P}$.
  - The probability mass function on X_t can be read off of q_t or, formally, $Pr(X_t = e_i) = q_t' e_i$.

4 Events and Observations

- The event space, Ω, is a set of all possible states of the world, ω.
  - An event, $\Lambda = \{\omega \mid .\} \subseteq \Omega$, is a set of possible states of the world, a subset of the event space.
  - The sigma algebra of measurable events, $\mathcal{F} = \{\Lambda \mid .\} \subset 2^{\Omega} = \{\Lambda \mid \Lambda \in \Omega\}$, is a set of 'most' events in the event space.
- The observation space, $\mathbb{R}^n$, is a set of possible observations, x.
  - An observation range, $b = \{x \mid .\} \subset \mathbb{R}^n$, is a set of possible observations, a subset of the observation space.
  - The sigma algebra of measurable observation ranges, $\mathcal{B} = \{b \mid .\} \subset 2^{\mathbb{R}^n} = \{b \mid b \in \mathbb{R}^n\}$ is a set of 'most' ranges in observation space.
- We may describe the event, Λ, that gives rise to any observation in range, b, by $\Lambda = \{\omega \mid X(\omega) \in b\}$
  - Describe the event $S^{-1}(\Lambda) := \{\omega \mid S(\omega) \in \Lambda\}$

5 Stationarity

- (Def.) A stochastic process, $(\{X_t\}_{t=0}^{\infty}, S, X, Pr)$, is stationary iff each element, X_t, has an identical associated unconditional probability measure, $Pr(X_t \in b) = Pr(X_0 \in b) \;\; \forall b \in \mathcal{B} \;\; \forall t$
- (Def.) A transformation, S, of a stochastic process, $(\{X_t\}_{t=0}^{\infty}, S, X, Pr)$, is measure preserving iff $Pr\{S^{-1}(\Lambda)\} = Pr\{\Lambda\} \;\; \forall \Lambda \in \mathcal{F}$.
- (Def.) A stationary measure, $Q : \mathcal{B} \to [0,1]$, of a Markov process, $(\{X_t\}_{t=0}^{\infty}, Q_0, P)$, solves the functional equation $\int P(dx^* \mid x)Q(dx) = Q(dx^*)$
- (Claim) A Markov process, $(\{X_t\}_{t=0}^{\infty}, Q_0, P)$, is stationary iff it is initiated at a stationary distribution, $Q_0 = Q$.
  - ($\Leftarrow$) $\forall t \forall b$, $Pr(X_{t+1} \in b) \overset{Pr(.)def.}{=} Q_{t+1}(b) \overset{Q_{t+1}def.}{=} \int P(b \mid x)Q_t(x) \overset{spse.Q_t=Q}{=} \int P(b \mid x)Q(x) \overset{Qdef.}{=} Q(b)$, $Q_0 = Q$, and induction.
  - ($\Rightarrow$) $\exists b$ s.t. $Pr(X_1 \in b) \overset{Pr(.)def.}{=} Q_1(b) \overset{Q_1def.}{=} \int P(b \mid x)Q_0(dx) \overset{Q_0 \neq Q}{\neq} Q_0(b) \overset{Pr(.)def.}{=} Pr(X_0 \in b) \Rightarrow\Leftarrow$.
- (Claim) A stochastic process, $(\{X_t\}_{t=0}^{\infty}, S, X, Pr)$, that represents a Markov process, $(\{X_t\}_{t=0}^{\infty}, Q_0, P)$, has a measure preserving transformation, S, iff the Markov process is initiated at a stationary measure, $Q_0 = Q$.

- ($\Rightarrow$) $\forall t \forall b$, $Pr(X_{t+1} \in b) \overset{Pr(.) def.}{=} Pr\{\omega \mid X_{t+1}(\omega) \in b\} \overset{X_{t+1} def.}{=} Pr\{\omega \mid X(S^{t+1}(\omega)) \in b\} \overset{\Lambda := \{\omega \mid X(S^t(\omega)) \in b\}}{=} Pr\{\omega \mid S(\omega) \in \Lambda\} \overset{S^{-1} def.}{=} Pr\{S^{-1}(\Lambda)\} \overset{S\, m.p.}{=} Pr\{\Lambda\} \overset{\Lambda def.}{=} Pr\{\omega \mid X(S^t(\omega)) \in b\} \overset{X_t def.}{=} Pr\{\omega \mid X_t(\omega) \in b\} \overset{Pr(.) def.}{=} Pr(X_t \in b)$ and induction.
- ($\Leftarrow$) $\forall t \forall b$, $Pr(X_{t+1} \in b) \overset{Pr(.) def.}{=} Q_{t+1}(b) \overset{Q sta.}{=} Q(b) \overset{Q sta.}{=} Q_0(b) \overset{Pr(.) def.}{=} Pr(X_0 \in b)$

- (Def.) A stationary vector, $q_{n\times 1}$ with $q'\underline{1} = 1$ and $q'e_i \geq 0\ \forall i$, of a discrete state Markov chain, $(\{X_t\}_{t=0}^{\infty}, q_0, \mathbb{P})$, solves the linear equation $q'\mathbb{P} = q'$.
- DSMC example?

6 Invariant Events and Unit Eigenfunctions

- (Def.) An event, Λ, of a stochastic process, $(\{X_t\}_{t=0}^{\infty}, S, X, Pr)$, is invariant iff $S^{-1}(\Lambda) = \Lambda$.
- (Def.) A unit eigenfunction, $\tilde{f}$, of a Markov Process, $(\{X_t\}_{t=0}^{\infty}, Q_0, P)$, satisfies the following (equivalent) functional equations: $\int \tilde{f}(x^*)P(dx^* \mid x) = \tilde{f}(x)$, $E[\tilde{f}(X_{t+1}) \mid X_t = x] = \tilde{f}(x)$, and $\mathbb{T}\tilde{f}(x) = \tilde{f}(x)$.
- (Claim) A function, $\tilde{f}$, is a unit eigenfunction of a Markov Process, $(\{X_t\}_{t=0}^{\infty}, Q_0, P)$, iff it is time-invariant, $\tilde{f}(X_t) = \tilde{f}(X_0)\ \forall t$, for some stationary measure, Q.
  - ($\Rightarrow$) $E[(\tilde{f}(X_{t+1}) - \tilde{f}(X_t))^2] \overset{alg.}{=} E[\tilde{f}(X_{t+1})^2] + E[\tilde{f}(X_t)^2] - 2E[\tilde{f}(X_{t+1})f(X_t)] \overset{LIE}{=} E[\tilde{f}(X_{t+1})^2] + E[\tilde{f}(X_t)^2] - 2E[E[\tilde{f}(X_{t+1}) \mid X_t]\tilde{f}(X_t)] \overset{\tilde{f} def.}{=} E[\tilde{f}(X_{t+1})^2] + E[\tilde{f}(X_t)^2] - 2E[\tilde{f}(X_t)\tilde{f}(X_t)] \overset{alg.}{=} E[\tilde{f}(X_{t+1})^2] - E[\tilde{f}(X_t)^2] \overset{\{X_t\}_{t=0}^{\infty} sta.}{=} 0$
  - ($\Leftarrow$) $E[\tilde{f}(X_{t+1}) \mid X_t] \overset{X_t = X_0\ \forall t}{=} E[\tilde{f}(X_t) \mid X_t] = \tilde{f}(X_t)$.
  - N.B. Though $\{X_t\}_{t=0}^{\infty}$ may vary over time, and though $\tilde{f}$ may vary over x, the ex-post realization of $\{\tilde{f}(X_t)\}_{t=0}^{\infty}$ is constant. The insight is that $\tilde{f}$ must be constant over any observation range, b, realizable in infinite time given X_0.
- (Claim) For a stochastic process, $(\{X_t\}_{t=0}^{\infty}, S, X, Pr)$, that describes a stationary Markov process, $(\{X_t\}_{t=0}^{\infty}, Q, P)$, an event is invariant, $S^{-1}(\Lambda) = \Lambda$, iff $\Lambda = \{\omega \mid \tilde{f}(X(\omega)) \in \tilde{b}\}$ for some unit eigenfunction of the Markov process, $\tilde{f}$, and range, $\tilde{b}$.
  - ($\Leftarrow$) $S^{-1}(\Lambda) \overset{\Lambda def.}{=} S^{-1}(\{\omega \mid \tilde{f}(X(\omega)) \in \tilde{b}\}) \overset{S^{-1} def}{=} \{\omega \mid \tilde{f}(X(S(\omega))) \in \tilde{b}\} \overset{X_1 def}{=} \{\omega \mid \tilde{f}(X_1(\omega)) \in \tilde{b}\} \overset{\tilde{f}(X_t) = \tilde{f}(X_0) \forall t}{=} \{\omega \mid \tilde{f}(X_0(\omega)) \in \tilde{b}\} \overset{X_0 def.}{=} \{\omega \mid \tilde{f}(X(\omega)) \in \tilde{b}\} \overset{\Lambda def}{=} \Lambda$

– (⇒) Claimed but not proved in Hansen's notes.

- (Def.) A unit eigenvector, v, of a discrete state Markov chain, $(\{X_t\}_{t=0}^{\infty}, q_0, \mathbb{P})$, solves the vector equation $\mathbb{P}v = v$.
- DSMC example?

7 Ergodicity

- (Def.) Ergodicity (?)
- (Def.) A transformation, S, of a stochastic process, $(\{X_t\}_{t=0}^{\infty}, S, X, Pr)$, is ergodic relative to its measure, Pr, iff invariant events are either certain or measure 0, $\forall \Lambda$ s.t. $S^{-1}(\Lambda) = \Lambda$, $Pr\{\Lambda\} \in \{0,1\}$.
- (Def.) A stationary Markov process, $(\{X_t\}_{t=0}^{\infty}, Q, P)$, is ergodic iff every unit eigenfunction, $\tilde{f}$, is constant wherever Q is non-zero, $\forall x$ s.t. $Q(dx) > 0$, $\tilde{f}(x) = c$.
- (Claim) A stationary Markov process, $(\{X_t\}_{t=0}^{\infty}, Q, P)$, is ergodic iff $\forall \tilde{f}\ \forall \tilde{b}$, $\int_{\{x|\tilde{f}(x)\in\tilde{b}\}} Q(dx) \in \{0,1\}$
  - (⇒) $\int_{\{x|\tilde{f}(x)\in\tilde{b}\}} Q(dx) \overset{Q def.}{=} \int_{\{x|\tilde{f}(x)\in\tilde{b},\, Q(dx)=0\}} Q(dx) + \int_{\{x|\tilde{f}(x)\in\tilde{b},\, Q(dx)>0\}} Q(dx)$ $\overset{\tilde{f}(x)=c\ \forall x\ s.t.\ Q(dx)>0}{=} 0 + \int_{\{x|c\in\tilde{b},\, Q(dx)>0\}} Q(dx) \overset{Q(X)=1;\ Q(\emptyset)=0}{=} \begin{cases} 1\ if\ c\in\tilde{b} \\ 0\ o/w \end{cases}$
  - (⇐) $(\{X_t\}_{t=0}^{\infty}, Q, P)$ not ergodic $\overset{erg.def.}{\Rightarrow} \exists(\tilde{f}, x_1, x_2)$ s.t. $Q(dx_1) > 0$, $Q(dx_2) > 0$, and $\tilde{f}(x_1) \neq \tilde{f}(x_2) \Rightarrow \exists(\tilde{f}, \tilde{b}, x_1, x_2)$ s.t. $Q(dx_1) > 0$, $Q(dx_2) > 0$, $\tilde{f}(x_1) \in \tilde{b}$, and $\tilde{f}(x_2) \notin \tilde{b} \Rightarrow 0 \overset{Q(dx_1)>0,\ \tilde{f}(x_1)\in\tilde{b}}{<} \int_{\{x|\tilde{f}(x)\in\tilde{b}\}} Q(dx) \overset{Q(X)=1}{=} 1 - \int_{\{x|\tilde{f}(x)\notin\tilde{b}\}} Q(dx) \overset{Q(dx_2)>0,\ \tilde{f}(x_2)\notin\tilde{b}}{<} 1 \Rightarrow\Leftarrow$
- (Claim) A stochastic process, $(\{X_t\}_{t=0}^{\infty}, S, X, Pr)$, that represents a stationary Markov process, $(\{X_t\}_{t=0}^{\infty}, Q, P)$, has an ergodic transformation, S, relative to its measure, Pr, iff the stationary Markov process is ergodic.
  - (Pf.) S ergodic relative to $Pr \overset{def.}{\Longleftrightarrow} \forall \Lambda \in \mathcal{J}$, $Pr\{\Lambda\} \in \{0,1\} \overset{\Lambda\in\mathcal{J}}{\Longleftrightarrow} \forall \tilde{f} \forall \tilde{b}$, $Pr\{\omega \mid \tilde{f}(X(\omega)) \in \tilde{b}\} \in \{0,1\} \overset{Pr(.)def.}{=} \forall \tilde{b}$, $Pr(X_t \in \{x \mid \tilde{f}(x) \in \tilde{b}\}) \in \{0,1\} \overset{Pr(.)def.}{\Longleftrightarrow} \forall \tilde{f} \forall \tilde{b}$, $\int_{\{x|\tilde{f}(x)\in\tilde{b}\}} Q(dx) \in \{0,1\} \overset{def.}{\Longleftrightarrow} (\{X_t\}_{t=0}^{\infty}, Q, P)$ is ergodic.
- (Def.) A stationary discrete state Markov chain, $(\{X_t\}_{t=0}^{\infty}, q, \mathbb{P})$, is ergodic iff given v, $\forall e_i$ s.t. $q' e_i > 0$, $e_i' v = c$.
- DSMC example?

8 Properties of Unit Eigenfunctions

- (Claim) If $\tilde{f}$ satisfies $\mathbb{T}\tilde{f} = \tilde{f}$, then $\tilde{f}$ satisfies $\mathbb{T}^j\tilde{f} = \tilde{f}\ \forall j \in \{1, 2, 3, ...\}$.
  - (Pf.) $\mathbb{T}^j\tilde{f} \overset{\mathbb{T}^j def.}{=} \mathbb{T}^{j-1}\mathbb{T}\tilde{f} \overset{\mathbb{T}\tilde{f}=\tilde{f}}{=} \mathbb{T}^{j-1}\tilde{f}$ and induction.
- (Claim) If $\tilde{f}$ satisfies $\mathbb{T}\tilde{f} = \tilde{f}$, then $\tilde{f}$ satisfies $\mathbb{M}\tilde{f} = \tilde{f}$ where $\mathbb{M} = (1 - \delta)\sum_{j=1}^{\infty}\delta^j\mathbb{T}^j$.
  - (Pf.) $\mathbb{M}\tilde{f} \overset{\mathbb{M}def.}{=} \left((1-\delta)\sum_{j=1}^{\infty}\delta^j\mathbb{T}^j\right)\tilde{f} \overset{alg.}{=} (1-\delta)\sum_{j=1}^{\infty}(\delta^j\mathbb{T}^j\tilde{f}) \overset{\mathbb{T}^j\tilde{f}=\tilde{f}}{=} (1-\delta)\sum_{j=1}^{\infty}(\delta^j\tilde{f}) \overset{alg.}{=} \left((1-\delta)\sum_{j=1}^{\infty}\delta^j\right)\tilde{f} = (1-\delta)\frac{1}{1-\delta}\tilde{f} = \tilde{f}$.
- (Claim) If $\tilde{f}$ satisfies $\mathbb{T}\tilde{f} = \tilde{f}$, then $f = \phi \circ \tilde{f}$ satisfies $\mathbb{T}f = f$.
  - (Pf.) $\mathbb{T}\tilde{f} = \tilde{f} \Rightarrow \tilde{f}(X_t) = \tilde{f}(X_0)\ \forall t \Rightarrow \phi \circ \tilde{f}(X_t) = \phi \circ \tilde{f}(X_0)\ \forall t \overset{f def.}{\Rightarrow} f(X_t) = f(X_0)\ \forall t \Rightarrow \mathbb{T}f = f$.
- (Claim) If $\tilde{f}$ satisfies $\mathbb{T}\tilde{f} = \tilde{f}$, then $f = \phi\circ\tilde{f}$ satisfies $\mathbb{T}^j f = f\ \forall j \in \{1, 2, 3, ...\}$ and $\mathbb{M}f = f$ where $\mathbb{M} = (1-\delta)\sum_{j=1}^{\infty}\delta^j\mathbb{T}^j$.
  - (Pf.) Combine the preceding results.

9 Sufficient Conditions for Markov Ergodicity

- (Claim) If $\forall f$ s.t. $f(x) \geq 0$ and $\int f(x)Q(dx) > 0$, $\mathbb{T}f(x) > 0\ \forall x$ s.t. $Q(dx) > 0$, then the stationary Markov process, $(\{X_t\}_{t=0}^{\infty}, Q, P)$, is ergodic, $\forall\tilde{f}\forall\tilde{b}$, $\int_{\{x|\tilde{f}(x)\in\tilde{b}\}}Q(dx) \in \{0, 1\}$.
  - (Pf.) $\int_{\{x|\tilde{f}(x)\in\tilde{b}\}}Q(dx) \neq 0 \overset{Q\geq 0}{\Rightarrow} \int_{\{x|\tilde{f}(x)\in\tilde{b}\}}Q(dx) > 0 \Rightarrow \exists x$ s.t. $Q(dx) > 0$ and $\tilde{f}(x) \in \tilde{b} \overset{f(x):=\mathbb{1}_{\{\tilde{f}(x)\in\tilde{b}\}}}{\Rightarrow} \exists x$ s.t. $Q(dx) > 0$ and $f(x) = 1 \overset{f(x)\geq 0\forall x, > for some x}{\Rightarrow} \int f(x)Q(dx) > 0 \overset{f\geq 0, \int f(x)Q(dx)>0}{\Rightarrow} \forall x$ s.t. $Q(dx) > 0$, $f(x) \overset{f=\phi o\tilde{f}}{=} \mathbb{T}f(x) \overset{ass.}{>} 0 \overset{f(x)\in\{0,1\}, f(x)\neq 0}{\Rightarrow} \forall x$ s.t. $Q(dx) > 0$, $f(x) = 1 \overset{f def.}{\Rightarrow} \forall x$ s.t. $Q(dx) > 0$, $\tilde{f}(x) \in \tilde{b} \Rightarrow \int_{\{x|\tilde{f}(x)\in\tilde{b}\}}Q(dx) \overset{Q def.}{=} 1$.
- (Claim) If $\forall f$ s.t. $f(x) \geq 0$ and $\int f(x)Q(dx) > 0$, $\exists j \in \{1, 2, 3, ...\}$ s.t. $\mathbb{T}^j f(x) > 0$, $\forall x$ s.t. $Q(dx) > 0$, then the stationary Markov process, $(\{X_t\}_{t=0}^{\infty}, Q, P)$, is ergodic.
  - (Pf.) Same as above using $f(x) := \mathbb{1}_{\{\tilde{f}(x)\in\tilde{b}\}} \Rightarrow f(x) = \mathbb{T}^j f(x)$.
- (Claim) If $\forall f$ s.t. $f(x) \geq 0$ and $\int f(x)Q(dx) > 0$, $\mathbb{M}f(x) > 0\ \forall x$ s.t. $Q(dx) > 0$, then the stationary Markov process, $(\{X_t\}_{t=0}^{\infty}, Q, P)$, is ergodic.
  - (Pf.) Same as above using $f(x) := \mathbb{1}_{\{\tilde{f}(x)\in\tilde{b}\}} \Rightarrow f(x) = \mathbb{M}f(x)$.
- DSMC example?

10 The Birkhoff LLN

- (Def.) For a stochastic process, $(\{X_t\}_{t=0}^{\infty}, S, X, Pr)$, the expected observation conditional on event Λ is given by $E[X \mid \Lambda] = \frac{\int_{\Lambda} X(\omega) Pr(d\omega)}{Pr(\Lambda)}$.
  - $E[X \mid \Lambda]$ is a description of the tendency of an observation $X(\omega)$ but not an 'empirical' one. It is computed based on information known ex-ante, Pr and Λ, not information revealed ex-post, $\{X_t(\omega)\}_{t=0}^{\infty}$.
  - The transformation, S, does not affect $E[X \mid \Lambda]$
- (Def.) For a stochastic process, $(\{X_t\}_{t=0}^{\infty}, S, X, Pr)$, we can compute the time-series average, $\frac{1}{N}\sum_{t=0}^{N} X_t$.
  - This is an 'empirical' description of the tendency of the observation, $X(\omega)$; it is computed based on information revealed ex-post, $\{X_t(\omega)\}_{t=0}^{\infty}$, not information known ex-ante, Pr and Λ.
  - The event space transformation, S, is responsible for generating the observations used in this computation.
- (Claim) For a stochastic process, $(\{X_t\}_{t=0}^{\infty}, S, X, Pr)$, with a measure preserving transformation, S, the time-series average converges to its expectation conditioned on invariant events, $\frac{1}{N}\sum_{t=0}^{N} X_t \overset{a.s.}{\to} E[X \mid \mathcal{J}]$.
  - With stationarity, the sample average converges to the conditional expectation, a r.v. over invariant events, not the unconditional expectation, the result in the standard LLN.
  - (Pf.) Due to Birkhoff
- (Claim) For a stochastic process, $(\{X_t\}_{t=0}^{\infty}, S, X, Pr)$, with a measure-preserving and ergodic transformation, S, the time-series average converges to the unconditional expectation, $\frac{1}{N}\sum_{t=0}^{N} X_t \overset{a.s.}{\to} E[X]$
  - With ergodicity, we fully recover the result of the standard LLN despite the dependence structure of a time-series sampling procedure.
  - (Pf.) $\frac{1}{N}\sum_{t=0}^{N} X_t \overset{a.s.}{\to} E[X \mid \mathcal{J}] \overset{S ergodic}{=} E[X \mid \Lambda \in \mathcal{J}] \overset{Pr\Lambda=1}{=} \int_{\Lambda \in \mathcal{J}} X(\omega) Pr(d\omega) \overset{\omega \in \Lambda^C \Rightarrow Pr(d\omega)=0}{=} \int_{\Lambda \cup \Lambda^C} X(\omega) Pr(d\omega) = E[X]$ (Ergodicity ensures that r.v. $E[X \mid \mathcal{J}]$ is degenerate and ALSO that the prevailing conditional expectation equals the unconditional expectation.)