Chapter 23

Ergodicity

Section 23.1 gives a general orientation to ergodic theory, which we will study in discrete time.

Section 23.2 introduces dynamical systems and their invariants, the setting in which we will prove our ergodic theorems.

Section 23.3 considers time averages, defines what we mean for a function to have an ergodic property (its time average converges), and derives some consequences.

Section 23.4 defines asymptotic mean stationarity, and shows that, with AMS dynamics, the limiting time average is equivalent to conditioning on the invariant sets.

23.1 General Remarks

To begin our study of ergodic theory, let us consider a famous[1] line from Gnedenko and Kolmogorov (1954, p. 1):

> In fact, all epistemological value of the theory of probability is based on this: that large-scale random phenomena in their collective action create strict, nonrandom regularity.

Now, this is how Gnedenko and Kolmogorov introduced their classic study of the limit laws for *independent* random variables, but most of the random phenomena we encounter around us are not independent. Ergodic theory is a study of how large-scale *dependent* random phenomena nonetheless create nonrandom regularity. The classical limit laws for IID variables $X_1, X_2, \ldots$ assert that, under the right conditions, sample averages converge on expectations,

$$\overline{X}_n \equiv \frac{1}{n}\sum_{i=1}^{n} X_i \rightarrow \mathbf{E}\left[X_i\right]$$

[1] Among mathematical scientists, anyway.

where the sense of convergence can be "almost sure" (strong law of large numbers), "L_p" (p^{th} mean), "in probability" (weak law), etc., depending on the hypotheses we put on the X_i. One meaning of this convergence is that sufficiently large random samples are representative of the entire population — that $\overline{X}_n$ makes a good estimate of $\mathbf{E}[X]$.

The ergodic theorems, likewise, assert that for *dependent* sequences $X_1, X_2, \ldots$, time averages converge on expectations

$$\overline{X}_t \equiv \frac{1}{t}\sum_{i=1}^{t} X_i \to \mathbf{E}[X_\infty]$$

where X_∞ is some limiting random variable, or in the most useful cases a *non*-random variable. Once again, the mode of convergence will depend on the kind of hypotheses we make about the random sequence X. Once again, the interpretation is that a *single* sample path is representative of the entire distribution over sample paths, *if* it goes on long enough.

Chapter 21 proved a mean-square (L_2) ergodic theorem for weakly stationary continuous-parameter processes. The next few chapters, by contrast, will develop ergodic theorems for non-stationary discrete-parameter processes.[2] This is a little unusual, compared to most probability books, so let me say a word or two about why. (1) Results we get will include stationary processes as special cases, but stationarity fails for many applications where ergodicity (in a suitable sense) holds. So this is more general and more broadly applicable. (2) Our results will all have continuous-time analogs, but the algebra is a lot cleaner in discrete time. (3) Some of the most important applications (for people like you!) are to statistical inference and learning with dependent samples, and to Markov chain Monte Carlo, and both of those are naturally discrete-parameter processes. We will, however, stick to continuous state spaces.

23.2 Dynamical Systems and Their Invariants

It is a very remarkable fact — but one with deep historical roots (von Plato, 1994, ch. 3) — that the way to get regular limits for stochastic processes is to first turn them into irregular deterministic dynamical systems, and then let averaging smooth away the irregularity. This section will begin by laying out dynamical systems, and their invariant sets and functions, which will be the foundation for what follows.

Definition 257 (Dynamical System) *A* dynamical system *consists of a measurable space Ξ, a σ-field $\mathcal{X}$ on Ξ, a probability measure μ defined on $\mathcal{X}$, and a mapping $T : \Xi \mapsto \Xi$ which is $\mathcal{X}/\mathcal{X}$-measurable.*

Remark: Measure-preserving transformations (Definition 53) are special cases of dynamical systems. Since (Theorem 52) every strongly stationary process can

[2] In doing so, I'm ripping off Gray (1988), especially chapters 6 and 7.

be represented by a measure-preserving transformation, namely the shift (Definition 48), the theory of ergodicity for dynamical systems which we'll develop is easily seen to include the usual ergodic theory of strictly-stationary processes as a special case. Thus, at the cost of going to the infinite-dimensional space of sample paths, we can always make it the case that the time-evolution is completely deterministic, and the only stochastic component to the process is its initial condition.

Lemma 258 (Dynamical Systems are Markov Processes) *Let $\Xi, \mathcal{X}, \mu, T$ be a dynamical system. Let $\mathcal{L}(X_1) = \mu$, and define $X_t = T^{t-1}X_1$. Then the X_t form a Markov process, with evolution operator K defined through $Kf(x) = f(Tx)$.*

Proof: For every $x \in \Xi$ and $B \in \mathcal{X}$, define $\kappa(x, B) \equiv \mathbf{1}_B(Tx)$. For fixed x, this is clearly a probability measure (specifically, the δ measure at Tx). For fixed B, this is a measurable function of x, because T is a measurable mapping. Hence, $\kappa(x, B)$ is a probability kernel. So, by Theorem 103, the X_t form a Markov process. By definition, $\mathbf{E}[f(X_1)|X_0 = x] = Kf(x)$. But the expectation is in this case just $f(Tx)$. □

Notice that, as a consequence, there is a corresponding operator, call it U, which takes signed measures (defined over $\mathcal{X}$) to signed measures, and specifically takes probability measures to probability measures.

Definition 259 (Observable) *A function $f : \Xi \mapsto \mathbb{R}$ which is $\mathbb{B}/\mathcal{X}$ measurable is an* observable *of the dynamical system $\Xi, \mathcal{X}, \mu, T$.*

Pretty much all of what follows would work if the observables took values in any real or complex vector space, but that situation can be built up from this one.

Definition 260 (Invariant Function, Invariant Set, Invariant Measure) *A function is invariant, under the action of a dynamical system, if $f(Tx) = f(x)$ for all $x \in \Xi$, or equivalently if $Kf = f$ everywhere. An event $B \in \mathcal{X}$ is invariant if its indicator function is an invariant function. A measure ν is invariant if it is preserved by T, i.e. if $\nu(C) = \nu(T^{-1}C)$ for all $C \in \mathcal{X}$, equivalently if $U\nu = \nu$.*

Lemma 261 *The class $\mathcal{I}$ of all measurable invariant sets in Ξ is a σ-algebra.*

Proof: Clearly, Ξ is invariant. The other properties of a σ-algebra follow because set-theoretic operations (union, complementation, etc.) commute with taking inverse images. □

Lemma 262 *An observable is invariant if and only if it is $\mathcal{I}$-measurable. Consequently, $\mathcal{I}$ is the σ-field generated by the invariant observables.*

PROOF: "If": Pick any Borel set B. Since $f = f \circ T$, $f^{-1}(B) = (f \circ T)^{-1}(B) = T^{-1}f^{-1}B$. Hence $f^{-1}(B) \in \mathcal{I}$. Since the inverse image of every Borel set is in $\mathcal{I}$, f is $\mathcal{I}$-measurable. "Only if": Again, pick any Borel set B. By assumption, $f^{-1}(B) \in \mathcal{I}$, so $f^{-1}(B) = T^{-1}f^{-1}(B) = (f \circ T)^{-1}(B)$, so the inverse image of under Tf of any Borel set is an invariant set, implying that $f \circ T$ is $\mathcal{I}$-measurable. Since, for every B, $f^{-1}(B) = (f \circ T)^{-1}(B)$, we must have $f \circ T = f$. The consequence follows. □

Definition 263 (Infinitely Often, i.o.) *For any set $C \in \mathcal{X}$, the set C* infinitely often*, $C_{\text{i.o.}}$, consists of all those points in Ξ whose trajectories visit C infinitely often, $C_{\text{i.o.}} \equiv \limsup_t T^{-t}C$.*

Lemma 264 *For every $C \in \mathcal{X}$, $C_{\text{i.o.}}$ is invariant.*

PROOF: Exercise. □

Definition 265 (Invariance Almost Everywhere) *A measurable function is* invariant μ-a.e., *or* almost invariant, *when*

$$\mu\{x \in \Xi | \forall n, \ f(x) = K^n f(x)\} = 1 \tag{23.1}$$

A measurable set is invariant μ-a.e., *when its indicator function is almost invariant.*

Remark 1: Some of the older literature annoyingly calls these objects *totally invariant*.

Remark 2: Invariance implies invariance μ-almost everywhere, for any μ.

Lemma 266 *The almost-invariant sets form a σ-field, $\mathcal{I}'$, and an observable is almost invariant if and only if it is measurable with respect to this field.*

PROOF: Entirely parallel to that for the strict invariants. □

Let's close this section with a simple lemma, which will however be useful in approximation-by-simple-function arguments in building up expectations.

Lemma 267 *A simple function, $f(x) = \sum_{k=1}^{m} a_m \mathbf{1}_{C_k}(x)$, is invariant if and only if all the sets $C_k \in \mathcal{I}$. Similarly, a*

simple function is almost invariant iff all the defining sets are almost invariant.

PROOF: Exercise. □

23.3 Time Averages and Ergodic Properties

For convenience, let's re-iterate the definition of a time average. (The notation differs here a little from that given earlier.)

Definition 268 (Time Average) *The time-average of an observable f is the real-valued function*

$$\overline{f}_t(x) \equiv \frac{1}{t}\sum_{i=0}^{t-1} f(T^i x) \quad (23.2)$$

The operator taking functions to their time-averages will be written $A_t f$:

$$A_t f(x) \equiv \overline{f}_t(x) \quad (23.3)$$

Lemma 269 *For every t, the time-average of an observable is an observable.*

PROOF: The class of measurable functions is closed under finite iterations of arithmetic operations. □

Definition 270 (Ergodic Property) *An observable f* has the ergodic property *when $\overline{f}_t(x)$ converges as $t \to \infty$ for μ-almost-all x. An observable* has the mean ergodic property *when $\overline{f}_t(x)$ converges in $L_1(\mu)$, and similarly for the other L_p* ergodic properties. *If for some class of functions $\mathcal{D}$, every $f \in \mathcal{D}$ has an ergodic property, then the class $\mathcal{D}$ has that ergodic property.*

Remark. Notice that what is required for f to have the ergodic property is that almost every initial point has *some* limit for its time average,

$$\mu\left\{x \in \Xi \middle| \exists r \in \mathbb{R} : \lim_{t\to\infty} \overline{f}_t(x) = r\right\} = 1 \quad (23.4)$$

Not that there is some *common* limit for almost every initial point,

$$\exists r \in \mathbb{R} : \mu\left\{x \in \Xi \middle| \lim_{t\to\infty} \overline{f}_t(x) = r\right\} = 1 \quad (23.5)$$

Similarly, a class of functions has the ergodic property if all of their time averages converge; they do not have to converge uniformly.

Definition 271 *If an observable f has the ergodic property, define $\overline{f}(x)$ to be the limit of $\overline{f}_t(x)$ where that exists, and* 0 *elsewhere. The corresponding operator will be written A:*

$$Af(x) = \overline{f}(x) \quad (23.6)$$

The domain of A consists of all and only the functions with ergodic properties.

Observe that

$$Af(x) = \lim_{t\to\infty} \frac{1}{t}\sum_{n=0}^{t} K^n f(x) \quad (23.7)$$

That is, A is the limit of an arithmetic mean of conditional expectations. This suggests that it should itself have many of the properties of conditional expectations. In fact, under a reasonable condition, we will see that $Af = \mathbf{E}[f|\mathcal{I}]$, expectation conditional on the σ-algebra of invariant sets. We'll check first that A has the properties we'd want from a conditional expectation.

Lemma 272 *A is a linear operator, and its domain is a linear space.*

PROOF: If c is any real number, then $A_t cf(x) = cA_t f(x)$, and so clearly, if the limit exists, $Acf(x) = cAf(x)$. Similarly, $A_t(f+g)(x) = A_t f(x) + A_t g(x)$, so if f and g both have ergodic properties, then so does $f+g$, and $A(f+g)(x) = Af(x) + Ag(x)$. □

Lemma 273 *If $f \in \mathrm{Dom}A$, and, for all $n \geq 0$, $fT^n \geq 0$ a.e., then $Af(x) \geq 0$ a.e.*

PROOF: The event $Af(x) < 0$ is a sub-event of $\bigcup_n \{f(T^n(x)) < 0\}$. Since the union of a countable collection of measure zero events has measure zero, $Af(x) \geq 0$ almost everywhere. □

We can't just say $f \geq 0$ a.e., because the effect of the transformation T might be to map every point to the bad set of f; the lemma guards against that. Of course, if $f(x) \geq 0$ for all, and not just almost all, x, then the bad set is non-existent, and $Af \geq 0$ follows automatically.

Lemma 274 *The constant function 1 has the ergodic property. Consequently, so does every other constant function.*

PROOF: For every n, $1(T^n x) = 1$. Hence $A_t 1(x) = 1$ for all t, and so $A1(x) = 1$. Extension to other constants follows by linearity. □

Remember that for any Markov operator K, $K1 = 1$.

Lemma 275 *If $f \in \mathrm{Dom}(A)$, then, for all n, $f \circ T^n$ is too, and $Af(x) = Af \circ T^n(x)$. Or, $AK^n f(x) = Af(x)$.*

PROOF: Start with $n = 1$, and show that the discrepancy goes to zero.

$$AKf(x) - Af(x) = \lim_t \frac{1}{t} \sum_{i=0}^{t} \left(K^{i+1} f(x) - K^i f(x)\right) \quad (23.8)$$

$$= \lim_t \frac{1}{t} \left(K^t f(x) - f(x)\right) \quad (23.9)$$

Since $Af(x)$ exists a.e., we know that the series $t^{-1}\sum_{i=0}^{t-1} K^i f(x)$ converges a.e., implying that $(t+1)^{-1} K^t f(x) \to 0$ a.e.. But $t^{-1} = \frac{t+1}{t}(t+1)^{-1}$, and for large t, $t + 1/t < 2$. Hence $(t+1)^{-1} K^t f(x) \leq t^{-1} K^t f(x) \leq 2(t+1)^{-1} K^t f(x)$, implying that $t^{-1} K^t f(x)$ itself goes to zero (a.e.). Similarly, $t^{-1} f(x)$ must go to zero. Thus, overall, we have $AKf(x) = Af(x)$ a.e., and $Kf(x) \in \mathrm{Dom}(A)$. □

Lemma 276 *If $f \in \mathrm{Dom}(A)$, then Af is an invariant, and $\mathcal{I}$-measurable.*

PROOF: Af exists, so (previous lemma) AKf exists and is equal to Af (almost everywhere). But $AKf(x) = Af(Tx)$, by definition, hence Af is invariant, i.e., $KAf = AKf = Af$. Measurability follows from Lemma 262. □

Lemma 277 *If $f \in \mathrm{Dom}(A)$, and B is any set in $\mathcal{I}$, then $A(\mathbf{1}_B(x)f(x)) = \mathbf{1}_B(x)Af(x)$.*

PROOF: For every n, $\mathbf{1}_B(T^n x)f(T^n x) = \mathbf{1}_B(x)f(T^n x)$, since $x \in B$ iff $T^n x \in B$. So, for all finite t, $A_t(\mathbf{1}_B(x)f(x)) = \mathbf{1}_B(x)A_t f(x)$, and the lemma follows by taking the limit. □

Lemma 278 *All indicator functions of measurable sets have ergodic properties if and only if all bounded observables have ergodic properties.*

PROOF: A standard approximation-by-simple-functions argument, as in the construction of Lebesgue integrals. □

Lemma 279 *Let f be bounded and have the ergodic property. Then Af is μ-integrable, and $\mathbf{E}[Af(X)] = \mathbf{E}[f(X)]$, where $\mathcal{L}(X) = \mu$.*

PROOF: Since f is bounded, it is integrable. Hence $A_t f$ is bounded, too, for all t, and $A_t f(X)$ is an integrable random variable. A sequence of bounded, integrable random variables is uniformly integrable. Uniform integrability, plus the convergence $A_t f(x) \rightarrow Af(x)$ for μ-almost-all x, gives us that $\mathbf{E}[Af(X)]$ exists and is equal to $\lim \mathbf{E}[A_t f(X)]$ via Fatou's lemma. (See e.g., Theorem 117 in the notes to 36-752.)

Now use the invariance of Af, i.e., the fact that $Af(X) = Af(TX)$ μ-a.s.

$$
\begin{aligned}
0 &= \mathbf{E}[Af(TX)] - \mathbf{E}[Af(X)] && (23.10)\\
&= \lim \frac{1}{t}\sum_{n=0}^{t-1}\mathbf{E}[K^n f(TX)] - \lim \frac{1}{t}\sum_{n=0}^{t-1}\mathbf{E}[K^n f(X)] && (23.11)\\
&= \lim \frac{1}{t}\sum_{n=0}^{t-1}\mathbf{E}[K^n f(TX)] - \mathbf{E}[K^n f(X)] && (23.12)\\
&= \lim \frac{1}{t}\sum_{n=0}^{t-1}\mathbf{E}\left[K^{n+1} f(X)\right] - \mathbf{E}[K^n f(X)] && (23.13)\\
&= \lim \frac{1}{t}\left(\mathbf{E}\left[K^t f(X)\right] - \mathbf{E}[f(X))\right] && (23.14)
\end{aligned}
$$

Hence

$$
\mathbf{E}[Af] = \lim \frac{1}{t}\sum_{n=0}^{t-1}\mathbf{E}[K^n f(X)] = \mathbf{E}[f(X)] \tag{23.15}
$$

as was to be shown. □

Lemma 280 *If f is as in Lemma 279, then $A_t f \rightarrow f$ in $L_1(\mu)$.*

PROOF: From Lemma 279, $\lim \mathbf{E}[A_t f(X)] = \mathbf{E}[f(X)]$. Since the variables $A_t f(X)$ are uniformly integrable (as we saw in the proof of that lemma), it follows (Proposition 4.12 in Kallenberg, p. 68) that they also converge in $L_1(\mu)$. □

Lemma 281 *Let f be as in Lemmas 279 and 280, and $B \in \mathcal{X}$ be an arbitrary measurable set. Then*

$$\lim_{t \to \infty} \frac{1}{t} \sum_{n=0}^{t-1} \mathbf{E}\left[\mathbf{1}_B(X) K^n f(X)\right] = \mathbf{E}\left[\mathbf{1}_B(X) f(X)\right] \tag{23.16}$$

where $\mathcal{L}(X) = \mu$.

PROOF: Let's write out the expectations explicitly as integrals.

$$\left| \int_B f(x) d\mu - \frac{1}{t} \sum_{n=0}^{t-1} \int_B K^n f(x) d\mu \right| \tag{23.17}$$

$$= \left| \int_B f(x) - \frac{1}{t} \sum_{n=0}^{t-1} K^n f(x) d\mu \right|$$

$$= \left| \int_B f(x) - A_t f(x) d\mu \right| \tag{23.18}$$

$$\leq \int_B \left| f(x) - A_t f(x) \right| d\mu \tag{23.19}$$

$$\leq \int \left| f(x) - A_t f(x) \right| d\mu \tag{23.20}$$

$$= \|f - A_t f\|_{L_1(\mu)} \tag{23.21}$$

But (previous lemma) these functions converge in $L_1(\mu)$, so the limit of the norm of their difference is zero. □

Boundedness is not essential.

Corollary 282 *Lemmas 279, 280 and 281 hold for any integrable observable $f \in \mathrm{Dom}(A)$, bounded or not, provided that $A_t f$ is a uniformly integrable sequence.*

PROOF: Examining the proofs shows that the boundedness of f was important only to establish the uniform integrability of $A_t f$. □

23.4 Asymptotic Mean Stationarity

Next, we come to an important concept which will prove to be necessary and sufficient for the most important ergodic properties to hold.

Definition 283 (Asymptotically Mean Stationary) *A dynamical system is* asymptotically mean stationary *when, for every $C \in \mathcal{X}$, the limit*

$$m(C) \equiv \lim_{t \to \infty} \frac{1}{t} \sum_{n=0}^{t-1} \mu(T^{-n} C) \tag{23.22}$$

exists, and the set function m is its stationary mean.

Remark 1: It might've been more logical to call this "asymptotically measure stationary", or something, but I didn't make up the names...

Remark 2: Symbolically, we can write

$$m = \lim_{t\to\infty} \frac{1}{t}\sum_{n=0}^{t-1} U^n \mu$$

where U is the operator taking measures to measures. This leads us to the next proposition.

Proposition 284 *If a dynamical system is stationary, i.e., T is preserves the measure μ, then it is asymptotically mean stationary, with stationary mean μ.*

PROOF: If T preserves μ, then for every measurable set, $\mu(C) = \mu(T^{-1}C)$. Hence every term in the sum in Eq. 23.22 is $\mu(C)$, and consequently the limit exists and is equal to $\mu(C)$. □

Theorem 285 (Vitali-Hahn Theorem) *If m_t are a sequence of probability measures on a common σ-algebra $\mathcal{X}$, and $m(C)$ is a set function such that $\lim_t m_t(C) = m(C)$ for all $C \in \mathcal{X}$, then m is a probability measure on $\mathcal{X}$.*

PROOF: This is a standard result from measure theory. □

Theorem 286 *If a dynamical system is asymptotically mean stationary, then its stationary mean is an invariant probability measure.*

PROOF: For every t, let $m_t(C) = \frac{1}{t}\sum_{n=0}^{t-1}\mu(T^{-n}(C))$. Then m_t is a linear combination of probability measures, hence a probability measure itself. Since, for every $C \in \mathcal{X}$, $\lim m_t(C) = m(C)$, by Definition 283, Proposition 285 says that $m(C)$ is also a probability measure. It remains to check invariance.

$$\begin{aligned}
& m(C) - m(T^{-1}C) && (23.23)\\
= \quad & \lim \frac{1}{t}\sum_{n=0}^{t-1}\mu(T^{-n}(C)) - \lim \frac{1}{t}\sum_{n=0}^{t-1}\mu(T^{-n}(T^{-1}C)) && \\
= \quad & \lim \frac{1}{t}\sum_{n=0}^{t-1}\mu(T^{-n-1}C) - \mu(T^{-n}C) && (23.24)\\
= \quad & \lim \frac{1}{t}\left(\mu(T^{-t}C) - \mu(C)\right) && (23.25)
\end{aligned}$$

Since the probability measure of any set is at most 1, the difference between two probabilities is at most 1, and so $m(C) = m(T^{-1}C)$, for all $C \in \mathcal{X}$. But this means that m is invariant under T (Definition 53). □

Remark: Returning to the symbolic manipulations, if μ is AMS with stationary mean m, then $Um = m$ (because m is invariant), and so we can write $\mu = m + (\mu - m)$, knowing that $\mu - m$ goes to zero under averaging. Speaking loosely (this can be made precise, at the cost of a fairly long excursion) m is an eigenvector of U (with eigenvalue 1), and $\mu - m$ lies in an orthogonal direction, along which U is contracting, so that, under averaging, it goes away, leaving only m, which is like the projection of the original measure μ on to the invariant manifold of U.

The relationship between an AMS measure μ and its stationary mean m is particularly simple on invariant sets: they are equal there. A slightly more general theorem is actually just as easy to prove, however, so we'll do that.

Lemma 287 *If μ is AMS with limit m, and f is an observable which is invariant μ-a.e., then $\mathbf{E}_{\mu}[f] = \mathbf{E}_{m}[f]$.*

PROOF: Let C be any almost invariant set. Then, for any t, C and $T^{-t}C$ differ by, at most, a set of μ-measure 0, so that $\mu(C) = \mu(T^{-t}C)$. The definition of the stationary mean (Equation 23.22) then gives $\mu(C) = m(C)$, or $\mathbf{E}_{\mu}[\mathbf{1}_C] = \mathbf{E}_{m}[\mathbf{1}_C]$, i.e., the result holds for indicator functions. By Lemma 267, this then extends to simple functions. The usual arguments then take us to all functions which are measurable with respect to $\mathcal{I}'$, the σ-field of almost-invariant sets, but this (Lemma 266) is the class of all almost-invariant functions. □

Lemma 288 *If μ is AMS with stationary mean m, and f is a bounded observable,*

$$\lim_{t\rightarrow\infty} \mathbf{E}_{\mu}[A_t f] = \mathbf{E}_{m}[f] \tag{23.26}$$

PROOF: By Eq. 23.22, this must hold when f is an indicator function. By the linearity of A_t and of expectations, it thus holds for simple functions, and so for general measurable functions, using boundedness to exchange limits and expectations where necessary. □

Lemma 289 *If f is a bounded observable in* Dom(A)*, and μ is AMS with stationary mean m, then $\mathbf{E}_{\mu}[Af] = \mathbf{E}_{m}[f]$.*

PROOF: From Lemma 281, $\mathbf{E}_{\mu}[Af] = \lim_{t\rightarrow\infty} \mathbf{E}_{\mu}[A_t f]$. From Lemma 288, the latter is $\mathbf{E}_{m}[f]$. □

Remark: Since Af is invariant, we've got $\mathbf{E}_{\mu}[Af] = \mathbf{E}_{m}[Af]$, from Lemma 287, but that's not the same.

Corollary 290 *Lemmas 288 and 289 continue to hold if f is not bounded, but $A_t f$ is uniformly integrable (μ).*

PROOF: As in Corollary 282. □

Theorem 291 *If μ is AMS, with stationary mean m, and the dynamics have ergodic properties for all the indicator functions, then, for any measurable set C,*

$$A\mathbf{1}_C = m(C|\mathcal{I}) \tag{23.27}$$

with probability 1 under both μ and m.

PROOF: By Lemma 276, $A\mathbf{1}_C$ is an invariant function. Pick any set $B \in \mathcal{I}$, so that $\mathbf{1}_B$ is also invariant. By Lemma 277, $A(\mathbf{1}_B\mathbf{1}_C) = \mathbf{1}_B A\mathbf{1}_C$, which is invariant (as a product of invariant functions). So Lemma 287 gives

$$\mathbf{E}_\mu\left[\mathbf{1}_B A\mathbf{1}_C\right] = \mathbf{E}_m\left[\mathbf{1}_B A\mathbf{1}_C\right] \tag{23.28}$$

while Lemma 289 says

$$\mathbf{E}_\mu\left[A(\mathbf{1}_B\mathbf{1}_C)\right] = \mathbf{E}_m\left[\mathbf{1}_B\mathbf{1}_C\right] \tag{23.29}$$

Since the left-hand sides are equal, the right-hand sides must be equal as well, so

$$\begin{aligned} m(B\cap C) &= \mathbf{E}_m\left[\mathbf{1}_B\mathbf{1}_C\right] && (23.30)\\ &= \mathbf{E}_m\left[\mathbf{1}_B A\mathbf{1}_C\right] && (23.31)\end{aligned}$$

Since this holds for all invariant sets $B \in \mathcal{I}$, we conclude that $A\mathbf{1}_C$ must be a version of the conditional probability $m(C|\mathcal{I})$. □

Corollary 292 *Under the assumptions of Theorem 291, for any bounded observable f,*

$$Af = \mathbf{E}_m\left[f|\mathcal{I}\right] \tag{23.32}$$

PROOF: From Lemma 278, every bounded observable has the ergodic property. One can then imitate the proof of the theorem to obtain the desired result. □

Corollary 293 *Equation 23.32 continues to hold if $A_t f$ are uniformly μ-integrable, or f is m-integrable.*

PROOF: Exercise. □