Breiman’s Law of Large Numbers

Itamar Cohen-Matalon

December 4, 2018

1 Introduction

This seminar will follow the proof of Breiman’s Law of Large Numbers from [Bre60] and [BQ16a]. We’ll begin with introducing the conceptually similar Birkhoff Ergodic Theorem (Compact Metric, Uniquely Ergodic case) with proof from [EW11].

2 Birkhoff - Uniquely Ergodic Case

Recall the Birkhoff Ergodic Theorem, and the Ergodic case:

Theorem 1 (Birkhoff Ergodic Theorem). Let $\left( X, \mathcal{B}, \mu, T \right)$ be a p.p.s., $f \in L^1 \left( X, \mathcal{B}, \mu \right)$. Then for $\mu$-a.e. $x \in X$:

$$f^* \left( x \right) = \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} f \left( T^n x \right)$$

The limit exists, $f^* \in L^1$, $f^* = f^* \circ T$. Denote by $\mathcal{E} \subseteq \mathcal{B}$ the set of $T$ invariant subsets. Then, $\forall A \in \mathcal{E}$:

$$\int_A f^* \, d\mu = \int_A f \, d\mu$$

Note: $f^* = E \left( f \mid \mathcal{E} \right)$

Corollary 1 (Pointwise Ergodic Theorem). If $T$ is ergodic, then $\mathcal{E}$ is trivial, and then $\forall f \in L^1$ and for $\mu$-a.e. $x \in X$:

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} f \left( T^n x \right) = \int f \, d\mu$$

Now we shall continue to an even more specific case, when the space $X$ is a compact metric space, and $T$ is uniquely ergodic (recall, a uniquely ergodic transformation is one that admits a single invariant probability measure).

Theorem 2 (Birkhoff - Uniquely Ergodic Case). Let $\left( X, \mathcal{B}, \mu, T \right)$ be a p.p.s., where $X$ is a compact metric space and $T$ is uniquely ergodic. The for every $x \in X$:

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} f \left( T^n x \right) = \int f \, d\mu$$

And this convergence is uniform across $X$.

We will need the following lemma (see appendix - 4.1 for proof):

Lemma 1. Let $\left( X, \mathcal{B}, \mu, T \right)$ be a p.p.s., where $X$ is a compact metric space. Then the space of probability measures: $\text{Prob} \left( X \right)$ is weak-* sequentially compact. It is characterized by:

$$\mu_n \to \mu \iff \forall f \in C \left( X \right) \int_X f \, d\mu_n \to \int_X f \, d\mu$$
We’ll start by proving another lemma:

**Lemma 2.** Let $T : X \to X$ be a continuous map of a compact metric space $X$, and $(\nu_n)$ be a sequence in $\text{Prob}(X)$. Then any weak-* limit point of the sequence $(\mu_n)$ defined by $\mu_n = \frac{1}{n} \sum_{j=0}^{n-1} T^j \nu_n$ is in $\text{Prob}^T(X)$ (the space of $T$-invariant probability measures).

Note: since by Lemma 1, $\text{Prob}^T(X)$ is weak-* compact, the sequence must have some limit point.

**Proof.** Notice: $\|f\|_\infty = \sup_{x \in X} |f| < \infty$ by compactness of $X$ and continuity of $f$.

Let $\mu_{n_k} \to \mu$ be a converging subsequence. We’d like to show that $\mu$ is $T$ invariant, i.e. $\forall f \in C(X)$:

$$\int_X f \, d\mu = \int_X f \circ T \, d\mu$$  \hspace{1cm} (6)

Recall, that by convergence we have:

$$\mu_{n_k} \to \mu \Rightarrow \forall g \in C(X) \int_X g \, d\mu_{n_k} \to \int_X g \, d\mu$$  \hspace{1cm} (7)

So we observe:

$$0 \leq \left| \int_X f \, d\mu_{n_k} - \int_X f \circ T \, d\mu_{n_k} \right| = \left| \frac{1}{n_k} \sum_{j=0}^{n_k-1} \int_X f \circ T^j \, d\nu_{n_k} - \frac{1}{n_k} \sum_{j=0}^{n_k-1} \int_X f \circ T^j \, d\nu_{n_k} \right| = \left| \frac{1}{n_k} \left( \int_X f \, d\nu_{n_k} - \int_X f \circ T \, d\nu_{n_k} \right) \right| \leq \frac{1}{n_k} \|f\|_\infty \to 0$$  \hspace{1cm} (8)

And so we must have that $\int_X f \, d\mu = \int_X f \circ T \, d\mu$ as required, so $\mu \in \text{Prob}^T(X)$. \hfill \Box

**Proof of Theorem 2.** Denote by $\mu$ the unique invariant measure. Take the sequence $(\delta_x)$ and apply Lemma 2 (i.e. in the notation of the lemma, we have: $\mu_n = \frac{1}{n} \sum_{j=0}^{n-1} \delta_{T^j x}$).

As the limit of every converging subsequence must be invariant (by the lemma) then by uniqueness it must be $\mu$. By compactness there are converging subsequences, and as the limit is unique we must have that:

$$\mu_n = \frac{1}{n} \sum_{j=0}^{n-1} \delta_{T^j x} \to \mu$$  \hspace{1cm} (9)

This is convergence in weak-* topology, so we have that for every $f \in C(X)$:

$$\frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) \to \int_X f \, d\mu$$  \hspace{1cm} (10)

And so we have everywhere convergence.

Now assume this convergence is not uniform. Denote: $E[f] = \int_X f \, d\mu$. So there is some $f \in C(X)$, $\epsilon > 0$ and $(N_k)$, such that $N_k \to \infty$ and for every $N_k$ exists $x_{N_k} \in X$ such that:

$$\left| \frac{1}{N_k} \sum_{j=0}^{N_k-1} f(T^j x_{N_k}) - E[f] \right| > \epsilon$$  \hspace{1cm} (11)

Taking the sequence $(\delta_{x_{N_k}})$ and in a similar manner to that of Lemma 2, we conclude that:

$$\mu_k = \frac{1}{N_k} \sum_{j=0}^{N_k-1} \delta_{T^j x_{N_k}} \to \mu$$  \hspace{1cm} (12)

And so:

$$\int_X f \, d\mu_k \to \int_X f \, d\mu = E[f]$$  \hspace{1cm} (13)
3 Breiman’s Law of Large Numbers

We will observe a Markov random walk on a compact metric space \((\Omega, \mathcal{B})\), determined by the transition probabilities, \(P(A \mid x)\) from \(x\) into \(A \in \mathcal{B}\), or alternatively defined by a Borel map from \(\Omega\) to to \(\text{Prob}(\Omega)\), \(x \mapsto P_x\).

We observe the space \((\Omega^\mathbb{N}, \mathcal{B}^\mathbb{N})\) the space of infinite sequences with coordinates in \(\Omega\), describing the random walk, and denote by \(X_n\) the random variable which describes the position at the \(n\)-th step, where \(X_0\) is the starting point of the random walk. For each \(x \in \Omega\), we define a probability \(\nu_x\) on \((\Omega^\mathbb{N}, \mathcal{B}^\mathbb{N})\) induced by \(P\) when \(X_0 = x\).

We shall consider only Markov-Feller processes: If \(T\) is the action defined on functions as:

\[
T \phi(x) = \int_\Omega \phi \, dP_x = E(\phi(X_1) \mid X_0 = x)
\]

Notice that \(T\) is a bounded operator w.r.t. the sup norm (as an operator on bounded functions): for every point, the value is an average of the original function over some set, and therefore not greater then the sup of the original function. Therefore, \(\|T(f)\|_\infty \leq \|f\|_\infty\), i.e. \(T\) is bounded. Note that as our space is compact metric, all continuous functions are bounded.

A Markov process is a Markov-Feller process if \(T\) acts on continuous functions (i.e. \(T(f)\) is continuous if \(f\) is). Note: This is equivalent to the map \(x \mapsto P_x\) being continuous w.r.t. the weak*- topology. Denote this map by \(M(x) = P_x\): In one direction, assume \(T\) acts on continuous functions. Notice that if \(x_n \to x\), then for every \(f \in C(\omega)\), as \(Tf\) is continuous:

\[
M(x_n)(f) = P_{x_n}(f) = \int_\Omega f \, dP_{x_n} = T \phi(x_n) \to T \phi(x) = \int_\Omega f \, dP_x = P_x(f) = M(x)(f)
\]

And so \(M\) is continuous w.r.t. to weak*- , as the integrals converge for every continuous \(f\).

Now assume that \(M\) is continuous. Then for every \(f \in C(\omega)\), \(x_n \to x\):

\[
T \phi(x_n) = \int_\Omega f \, dP_{x_n} = P_{x_n}(f) = M(x_n)(f) \to M(x)(f) = P_x(f) = \int_\Omega f \, dP_x = T \phi(x)
\]

And so \(Tf\) is continuous as required.

Example 1. (Group Action) Consider the case where we have a random walk on a compact metric space \((\Omega, \mathcal{B})\), determined by a (continuous) group action of \(G\) on \(\Omega\), where choice of \(G\) action is determined by a compactly supported probability measure on \(G, \mu_G\).

This is a Markov-Feller operator, as we can see that for \(\phi \in C(\Omega)\):

\[
T \phi(x) = \int_\Omega \phi \, dP_x = \int_G \phi(gx) \mu_G(g)
\]

Here we have that \(P_x\) is the push forward of \(\mu_G\) under the map \(g \mapsto gx\).

Notice that because \(\Omega\) is compact, \(\phi\) is equicontinuous. Choose some \(\epsilon > 0\), and \(\epsilon_1 = \epsilon(\epsilon) > 0\) that exists by equicontinuity of \(\phi\) s.t. if \(|x - y| < \epsilon_1\), then \(|\phi(x) - \phi(y)| < \epsilon\). Each \(g\) is also equicontinuous, with a corresponding \(\epsilon_g\) s.t. if \(|x - y| < \epsilon_g\) then \(|gx - gy| < \epsilon_g|\). For each \(g\) we can take the neighborhood where the difference is \(< \frac{\epsilon}{2}\) by the sup norm, and this gives an open cover, so by compactness we have a finite subcover. Note that for \(g\) we have that if \(h\) is in it’s neighborhood, then if \(|x - y| < \epsilon_g\) then \(|gx - gy| < \epsilon_1\).
Proposition 1. Observe a Markov-Feller random walk on a compact metric base space always exists.

In the setting of a Markov-Feller random walk on a compact metric base space weak-* limit point of the sequence \( (g) \) gives us that \( |T\phi(y) - T\phi(x)| < \epsilon \), i.e. we have continuity.

**Proof.** Proof is somewhat similar to that of Lemma 2.

We shall see that in our setting, a stationary measure always exists. Recall we have the operator \( T : C \rightarrow C \) defined by:

\[
T \phi(x) = \int_{\Omega} \phi \ dP_x = E (\phi(X_1) \mid X_0 = x)
\]

We have: \( (T^k\phi)(x) = E (\phi(X_k) \mid X_0 = x) \). We’ve shown that \( T \) is bounded, with \( \|Tf\|_\infty \leq \|f\|_\infty \), so \( \|T^n f\|_\infty = \|f\|_\infty \) too.

In addition denote: \( T_N \phi = \frac{1}{N} \sum_{n=1}^{N} T^n \phi \).

**Proposition 1.** Observe a Markov-Feller random walk on \( \Omega \). Let \( (\nu_n) \) be a sequence in Prob \( \Omega \). Then any weak-\(^*\) limit point of the sequence \( (\mu_n) \) defined by \( \mu_n = \frac{1}{n} \sum_{j=0}^{n-1} T_j \nu_n \) is stationary.

**Proof.** Proof is somewhat similar to that of Lemma 2.

Denote a limit point of a subsequence by \( \mu \). We would like to show that is is a stationary measure, i.e. that for all measurable \( A \):

\[
\mu(A) = \int_{\Omega} P(A \mid x) \ d\mu = \int_{\Omega} E (1_A(X_1) \mid X_0 = x) \ d\mu = (T_\mu)(A)
\]

So we would like to show that \( \mu = T_\mu \). It would suffice to prove for each \( f \in C(\Omega) \) that:

\[
\int_{\Omega} f \ d\mu = \int_{\Omega} f \ d(T_\mu) = \int_{\Omega} T f \ d\mu
\]

But as \( f \in C(\Omega) \) we have that \( T f \in C(\Omega) \), so by weak-\(^*\) convergence of the subsequence \( (\mu_{n_k}) \), we have:

\[
\int_{\Omega} f \ d\mu_{n_k} \rightarrow \int_{\Omega} f \ d\mu \quad \text{and} \quad \int_{\Omega} T f \ d\mu_{n_k} \rightarrow \int_{\Omega} T f \ d\mu
\]

Observe that:

\[
0 \leq \left| \int_{\Omega} f \ d\mu_{n_k} - \int_{\Omega} T f \ d\mu_{n_k} \right| = \left| \frac{1}{n_k} \sum_{j=0}^{n_k-1} \int_{\Omega} T_j f \ d\nu_{n_k} - \frac{1}{n_k} \sum_{j=0}^{n_k-1} \int_{\Omega} T_j f \ d\nu_{n_k} \right| = \\
\left| \frac{1}{n_k} \left( \int_{\Omega} f \ d\nu_{n_k} - \int_{\Omega} T^{n_k} f \ d\nu_{n_k} \right) \right| \leq \frac{1}{n_k} (\|f\|_\infty + \|T^n f\|_\infty) \leq \frac{2}{n_k} \|f\|_\infty \rightarrow 0
\]

And therefore we must have that the limits are equal as required, and \( \mu \) is stationary.

**Corollary 2.** In the setting of a Markov-Feller random walk on a compact metric base space \( (\Omega, B) \), a stationary measure always exists.
Proof. Take an arbitrary sequence of measures \((\nu_n)\) in Prob \((\Omega)\), and observe the sequence \((\mu_n)\) as defined in Proposition 1. By Lemma 1, Prob \((\Omega)\) is sequentially compact, so \((\mu_n)\) has a converging subsequence to some \(\mu\). By Proposition 1 this limit is stationary, and so a stationary probability measure exists, as required.

From here on, we shall assume that there exists a **unique** stationary measure.

**Theorem 3** (Breiman’s Law of Large Numbers). In the setting of a Markov-Feller random walk on a compact metric base space \((\Omega, B)\) with a unique stationary measure \(\pi\), we have that \(\forall \phi \in C(\Omega)\), then for any \(x \in \Omega\) for \(\nu_x\) a.e. sequence \((X_N)\):

\[
\frac{1}{N} \sum_{n=1}^{N} \phi(X_n) \to \int_{\Omega} \phi(x) \, d\pi(x) \tag{27}
\]

Note the similarities to the Birkhoff theorem: the time averages converge to the space average, for any starting point, and almost every path.

**Proposition 2.** For any \(\phi \in C(\Omega)\), \(\bar{T}_N(\phi)\) converges uniformly to \(\pi(\phi) = \int_{\Omega} \phi \, d\pi\).

**Proof.** Proof is somewhat similar to that of Theorem 2. Let \((\nu_n)\) be a sequence in Prob \((\Omega)\). Then any weak-* limit point of the sequence \((\mu_n)\) defined by \(\mu_n = \frac{1}{N} \sum_{j=0}^{N-1} T^j \nu_n\) is stationary by Proposition 1, and therefore \(\pi\). By compactness, this is the limit of the entire sequence \((\mu_n)\).

As \(\phi\) is continuous, we get that taking \(\nu^\epsilon_x = \delta_x\) gives us everywhere convergence, i.e.:

\[
\bar{T}_n(\phi)(x) = \int_{\Omega} \phi \, d\mu^\epsilon_n \to \int_{\Omega} \phi \, d\pi \tag{28}
\]

Now assume this convergence is not uniform. Denote: \(E[\phi] = \int_{\Omega} \phi \, d\mu\). So there is some \(f \in C(\Omega), \epsilon > 0\) and \((N_k)\), such that \(N_k \to \infty\) and for every \(N_k\) exists \(x_{N_k} \in \Omega\) such that:

\[
\left| \frac{1}{N_k} \sum_{j=0}^{N_k-1} T^j f(x_{N_k}) - E[f] \right| > \epsilon \tag{29}
\]

Taking the sequence \(\left(\delta_{x_{N_k}}\right)\) we get:

\[
\mu_k = \frac{1}{N_k} \sum_{j=0}^{N_k-1} T^j \delta_{x_{N_k}} \to \pi \tag{30}
\]

And so:

\[
\int_{\Omega} f \, d\mu_k \to \int_{\Omega} f \, d\pi = E[f] \tag{31}
\]

But by (29), notice that for all \(N_k\):

\[
\left| \int_{\Omega} f \, d\mu_k - E[f] \right| = \left| \frac{1}{N_k} \sum_{j=0}^{N_k-1} T^j f(x_{N_k}) - E[f] \right| > \epsilon \tag{32}
\]

And so the limit must be \(\geq \epsilon\) - a contradiction. Therefore we must have uniform convergence.

The following lemma is a derivative of Doob’s martingale convergence theorem, and a proof can be found in 4.2.

**Lemma 3.** Let \((Y_n)_{n \geq 1}\) be a sequence of random variables which are uniformly bounded in \(L^2\) (i.e. \(\exists M < \infty\) s.t. for all \(n \geq 1\), \(E(Y_n^2) < M\)), and such that:

\[
E(Y_n \mid Y_1, \ldots, Y_{n-1}) = 0 \quad \forall n \geq 1 \tag{33}
\]

Then the sequence \(\frac{1}{n} \sum_{k=1}^{n} Y_k\) converges to 0 almost surely and in \(L^2\).

5
Proposition 3. Let $\phi \in C(\Omega)$, $x \in \Omega$. Let $(X_n)_{n \geq 1}$ be distributed according to $\nu_x$ and define:

\[
Z_n^1 = \begin{cases} 
\phi(X_n) - E(\phi(X_n) \mid X_{n-1}) & n > 1 \\
0 & n \leq 1
\end{cases}
\]

\[
Z_n^k = \begin{cases} 
E(\phi(X_n) \mid X_{n-k+1}) - E(\phi(X_n) \mid X_{n-k}) & n > k \\
0 & n \leq k
\end{cases}
\]  

(34)

Then $\frac{1}{N} \sum_{n=1}^{N} Z_n^k \to 0$ a.s. $\nu_x$.

Proof. To prove we would like to use Lemma 3. First note that clearly as $Z_n^k \leq 2 \|\phi\|_{\infty}$, we have that $E\left(\left(Z_n^k\right)^2\right) \leq 4 \|\phi\|_{\infty}^2 < \infty$ for all $n, k$.

Next we must check that: $E(Z_n^k \mid Z_{n-1}^k, ..., Z_1^k) = 0$. Note that $Z_{n-1}^k, ..., Z_1^k$ are random variables in the $\sigma$-algebra of $X_{n-k}, ..., X_1$, so:

\[
E(Z_n^k \mid Z_{n-1}^k, ..., Z_1^k) = E(E(Z_n^k \mid X_{n-k}, ..., X_1) \mid Z_{n-1}^k, ..., Z_1^k)
\]  

(35)

Notice that by the Markov property:

\[
E(Z_n^k \mid X_{n-k}, ..., X_1) = E(Z_n^k \mid X_{n-k}) = E(E(\phi(X_n) \mid X_{n-k+1}) \mid X_{n-k}) - E(\phi(X_n) \mid X_{n-k}) = E(E(\phi(X_n) \mid X_{n-k+1}, X_{n-k}) \mid X_{n-k}) - E(\phi(X_n) \mid X_{n-k}) = E(\phi(X_n) \mid X_{n-k}) - E(\phi(X_n) \mid X_{n-k}) = 0
\]  

(36)

And so we have all that is required in order to use Lemma 3 which gives us the desired result. \square

Proof. Breiman’s Law of Large Numbers

To prove the theorem, we first notice that for $n > k$:

\[
\phi(X_n) - E(\phi(X_n) \mid X_{n-k}) = \sum_{i=1}^{k} Z_i^n
\]  

(37)

By Proposition 3 we have that for $(X_n)$ a.e.-$\nu_x$ (notice that we are neglecting $k$ bounded terms when comparing to $\sum_{i=1}^{k} \sum_{n=1}^{N} Z_i^n$, but in the limit these are negligible - as we take $k$ to be constant, and we have the $\frac{1}{N}$ normalization):

\[
\frac{1}{N} \sum_{n=1}^{N} \phi(X_n) - E(\phi(X_n) \mid X_{n-k}) \to 0
\]  

(38)

Again neglecting at most $k$ bounded (by $\|\phi\|_{\infty}$) terms, we get:

\[
\frac{1}{N} \sum_{n=1}^{N} \phi(X_n) - \frac{1}{N} \sum_{n=1}^{N} E(\phi(X_{n+k}) \mid X_n) \to 0
\]  

(39)

Summing up this limit for $k \leq M$ for some fixed $M$ (and dividing by $M$) we get get:

\[
\frac{1}{M} \sum_{n=1}^{N} \phi(X_n) - \frac{1}{M} \sum_{n=1}^{N} \frac{1}{M} \sum_{k=1}^{M} E(\phi(X_{n+k}) \mid X_n) \to 0
\]  

(40)

By Proposition 3 for any $\epsilon > 0$, we can choose an $M$, s.t. for all $n$:

\[
\left| \frac{1}{M} \sum_{k=1}^{M} E(\phi(X_{n+k}) \mid X_n) - \pi(\phi) \right| < \epsilon
\]  

(41)
As we have uniform convergence.

For such a choice of $M$, we get via substituting in (40) that $\nu_x$-a.e.:

$$\lim_{N} \frac{1}{N} \left| \sum_{n=1}^{N} \phi(X_n) - \pi(\phi) \right| \leq \epsilon$$  \hspace{1cm} (42)

As $\epsilon$ is arbitrary, taking a countable sequence going to 0 will give us the required result, that $\nu_x$-a.e.:

$$\lim_{N} \frac{1}{N} \left| \sum_{n=1}^{N} \phi(X_n) - \pi(\phi) \right| = 0$$  \hspace{1cm} (43)

\[ \square \]

4 Appendix

4.1 Proof of Lemma 1

The space of signed measures (linear functionals), $M$ is the dual of $C_0(X)$. As $X$ is compact, we have $C_0(X) = C_b(X) = C(X)$. By Banach-Alaoglu, the unit ball $B$ in $M$ is sequentially compact w.r.t. the weak-* topology. Notice that $\text{Prob}(X) \subseteq B$, and that it is closed, as if $\mu_n \to \mu$ for $\mu_n \in \text{Prob}(X)$, then $\mu$ must still be a positive functional (as $\mu_n$ are), and as $1_X \in C_b(X) = C_0(X)$, we have that:

$$1 = \mu_n(X) = \langle 1_X, \mu_n \rangle \to \langle 1_X, \mu \rangle = \mu(X)$$  \hspace{1cm} (44)

And so $\mu(X) = 1$ (and is a positive functional), and therefore $\mu \in \text{Prob}(X)$. Therefore $\text{Prob}(X)$ is a closed subset of a sequentially compact set (the unit ball in $M$), and is therefore sequentially compact w.r.t. the weak-* topology.

4.2 Proof of Lemma 3

See [BQ16b], Corollary A.8.

References


