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 $(X, \mathcal{B}, \mu, T)$ be a p.p.s., $f \in L^1(X, \mathcal{B}, \mu)$. Then for μ-a.e. $x \in X$:*

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

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, $\forall A \in \mathcal{E}$:

$$\int_A f^* \, \mathrm{d}\mu = \int_A f \, \mathrm{d}\mu \tag{2}$$

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

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

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) = \int f \, \mathrm{d}\mu \tag{3}$$

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 $(X, \mathcal{B}, \mu, T)$ be a p.p.s., where X is a compact metric space and T is uniquely ergodic. The for **every** $x \in X$:*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) = \int f \, \mathrm{d}\mu \tag{4}$$

*And this convergence is **uniform** across X.*

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

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

$$\mu_n \to \mu \Leftrightarrow \forall f \in C(X) \int_X f \, \mathrm{d}\mu_n \to \int_X f \, \mathrm{d}\mu \tag{5}$$

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 (ν_n) be a sequence in* $\mathrm{Prob}(X)$*. Then any weak-* limit point of the sequence (μ_n) defined by $\mu_n = \frac{1}{n}\sum_{j=0}^{n-1} T_*^j \nu_n$ is in* $\mathrm{Prob}^T(X)$ *(the space of T invariant probability measures).*

Note: since by Lemma 1, $\mathrm{Prob}^T(X)$ *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 μ is T invariant, i.e. $\forall f \in C(X)$:

$$\int_X f \,\mathrm{d}\mu = \int_X f \circ T \,\mathrm{d}\mu \tag{6}$$

Recall, that by convergence we have:

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

So we observe:

$$0 \le \left|\int_X f \,\mathrm{d}\mu_{n_k} - \int_X f\circ T \,\mathrm{d}\mu_{n_k}\right| = \left|\frac{1}{n_k}\sum_{j=0}^{n_k-1}\int_X f\circ T^{j-1}\,\mathrm{d}\nu_{n_k} - \frac{1}{n_k}\sum_{j=0}^{n_k-1}\int_X f\circ T^{j}\,\mathrm{d}\nu_{n_k}\right| =$$
$$\left|\frac{1}{n_k}\left(\int_X f\,\mathrm{d}\nu_{n_k} - \int_X f\circ T^{n_k}\,\mathrm{d}\nu_{n_k}\right)\right| \le \frac{1}{n_k}2\|f\|_\infty \to 0 \tag{8}$$

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

Proof of Theorem 2. Denote by μ the unique invariant measure. Take the sequence (δ_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 μ. By compactness there are converging subsequences, and as the limit is unique we must have that:

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

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

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

And so we have everywhere convergence.

Now assume this convergence is not uniform. Denote: $E[f] = \int_X f \,\mathrm{d}\mu$. 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:

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

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

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

And so:

$$\int_X f \,\mathrm{d}\mu_k \to \int_X f \,\mathrm{d}\mu = E[f] \tag{13}$$

But by (11), notice that for all N_k:

$$\left|\int_X f \, \mathrm{d}\mu_k - E[F]\right| = \left|\frac{1}{N_k}\sum_{j=0}^{N_k-1} f\left(T^j x_{N_k}\right) - E[F]\right| > \epsilon \tag{14}$$

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

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 Ω to to $\mathrm{Prob}(\Omega)$, $x \mapsto P_x$.

We observe the space $\left(\Omega^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}\right)$ the space of infinite sequences with coordinates in Ω, 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 ν_x on $\left(\Omega^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}\right)$ induced by P when $X_0 = x$.

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

$$T\phi(x) = \int_\Omega \phi \, \mathrm{d}P_x = E(\phi(X_1) \mid X_0 = x) \tag{15}$$

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:

$$M(x_n)(f) = P_{x_n}(f) = \int_\Omega f \, \mathrm{d}P_{x_n} = T\phi(x_n) \to T\phi(x) = \int_\Omega f \, \mathrm{d}P_x = P_x(f) = M(x)(f) \tag{16}$$

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 $f \in C(\omega)$, $x_n \to x$:

$$T\phi(x_n) = \int_\Omega f \, \mathrm{d}P_{x_n} = P_{x_n}(f) = M(x_n)(f) \to M(x)(f) = P_x(f) = \int_\Omega f \, \mathrm{d}P_x = T\phi(x) \tag{17}$$

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 Ω, where choice of G action is determined by a compactly supported probability measure on G, μ_G.

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

$$T\phi(x) = \int_\Omega \phi \, \mathrm{d}P_x = \int_G \phi(gx) \, \mu_G(g) \tag{18}$$

Here we have that P_x is the push forward of μ_G under the map $g \mapsto gx$.

Notice that because Ω is compact, ϕ is equicontinuous. Choose some $\epsilon > 0$, and $\epsilon_1 = \epsilon(\epsilon) > 0$ that exists by equicontinuity of ϕ s.t. if $|x - y| < \epsilon_1$, then $|\phi(x) - \phi(y)| < \epsilon$. Each g is also equicontinuous, with a corresponding ϵ_g s.t. if $|x - y| < \epsilon_g$ then $|gx - gy| < \frac{\epsilon_1}{2}$. For each g we can take the neighborhood where the difference is $< \frac{\epsilon_1}{4}$ 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 $|gx - gy| < \epsilon_1$.

Taking the minimum over the ϵ_g for the g's chosen for the subcover, we get some $\epsilon_2 > 0$ such that for all g, if $|x-y| < \epsilon_2$ then $|gx - gy| < \epsilon_1$. And so $|\phi(gx) - \phi(gy)| < \epsilon$, so integrating (against the probability measure μ_G) gives us that $|T\phi(y) - T\phi(x)| < \epsilon$ - i.e. we have continuity.

Definition 1. (Stationary Measure) A measure π on Ω shall be called stationary if $\forall A \in \mathcal{B}$:

$$\pi(A) = \int_\Omega P(A \mid x) \, \mathrm{d}\pi(x) \tag{19}$$

Note that in Example 1, we have that this translates to: $\forall A \in \mathcal{B}$:

$$\pi(A) = \int_\Omega P(A \mid x) \, \mathrm{d}\pi(x) = \int_G \int_\Omega \mathbb{1}_A(gx) \, \mathrm{d}\pi(x) \, \mathrm{d}\mu_G(g) = \mu_G * \pi(A) \tag{20}$$

I.e., π is stationary if $\pi = \mu_G * \pi$, where $\mu_G * \pi$ is convolution of the two measures, which is defined by:

$$\mu_G * \pi(A) = \int_G \int_\Omega \mathbb{1}_A(gx) \, \mathrm{d}\pi(x) \, \mathrm{d}\mu_G(g) \tag{21}$$

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

$$T\phi(x) = \int_\Omega \phi \, \mathrm{d}P_x = E(\phi(X_1) \mid X_0 = x) \tag{22}$$

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

In addition denote: $\bar{T}_N\phi = \frac{1}{N}\sum_{n=1}^N T^n\phi$.

Proposition 1. *Observe a Markov-Feller random walk on Ω. Let (ν_n) be a sequence in* $\mathrm{Prob}(\Omega)$*. Then any weak-* limit point of the sequence (μ_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 μ. We would like to show that is is a stationary measure, i.e. that for all measurable A:

$$\mu(A) = \int_\Omega P(A \mid x) \, \mathrm{d}\mu = \int_\Omega E(\mathbb{1}_A(X_1) \mid X_0 = x) \, \mathrm{d}\mu = (T_*\mu)(A) \tag{23}$$

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

$$\int_\Omega f \, \mathrm{d}\mu \underset{(?)}{=} \int_\Omega f \, \mathrm{d}(T_*\mu) = \int_\Omega Tf \, \mathrm{d}\mu \tag{24}$$

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

$$\int_\Omega f \, \mathrm{d}\mu_{n_k} \to \int_\Omega f \, \mathrm{d}\mu \qquad \text{and} \qquad \int_\Omega Tf \, \mathrm{d}\mu_{n_k} \to \int_\Omega Tf \, \mathrm{d}\mu \tag{25}$$

Observe that that:

$$0 \leq \left| \int_\Omega f \, \mathrm{d}\mu_{n_k} - \int_\Omega Tf \, \mathrm{d}\mu_{n_k} \right| = \left| \frac{1}{n_k}\sum_{j=0}^{n_k-1} \int_\Omega T^{j-1} f \, \mathrm{d}\nu_{n_k} - \frac{1}{n_k}\sum_{j=0}^{n_k-1} \int_\Omega T^j f \, \mathrm{d}\nu_{n_k} \right| =$$

$$\left| \frac{1}{n_k}\left(\int_\Omega f \, \mathrm{d}\nu_{n_k} - \int_\Omega T^{n_k} f \, \mathrm{d}\nu_{n_k} \right) \right| \leq \frac{1}{n_k}\left(\|f\|_\infty + \left\|T^k f\right\|_\infty \right) \leq \frac{2}{n_k}\|f\|_\infty \to 0 \tag{26}$$

And therefore we must have that the limits are equal as required, and μ is stationary. □

Corollary 2. *In the setting of a Markov-Feller random walk on a compact metric base space $(\Omega, \mathcal{B})$, a stationary measure always exists.*

Proof. Take an arbitrary sequence of measures (ν_n) in $\mathrm{Prob}(\Omega)$, and observe the sequence (μ_n) as defined in Proposition 1. By Lemma 1, $\mathrm{Prob}(\Omega)$ is sequentially compact, so (μ_n) has a converging subsequence to some μ. By Proposition 1 this limit is stationary, and so a stationary probability measure exists, as required. $\square$

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 $(\Omega, \mathcal{B})$ with a unique stationary measure π, we have that $\forall \phi \in C(\Omega)$, then for any $x \in \Omega$ for ν_x a.e. sequence (X_N):*

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

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 $\pi(\phi) = \int_{\Omega}\phi\ \mathrm{d}\pi$.*

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

As ϕ is continuous, we get that taking $\nu_n^x = \delta_x$ gives us everywhere convergence, i.e.:

$$\bar{T}_n(\phi)(x) = \int_{\Omega}\phi\ \mathrm{d}\mu_n^x \to \int_{\Omega}\phi\ \mathrm{d}\pi \tag{28}$$

Now assume this convergence is not uniform. Denote: $E[\phi] = \int_{\omega}\phi\ \mathrm{d}\mu$. 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:

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

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

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

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

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 $n \geq 1$, $E\left(Y_n^2\right) < M$), and such that:*

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

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

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

$$\begin{aligned} Z_n^1 &= \begin{cases} \phi(X_n) - E(\phi(X_n) \mid X_{n-1}) & n > 1 \\ 0 & n \leq 1 \end{cases} \\ Z_n^k &= \begin{cases} E(\phi(X_n) \mid X_{n-k+1}) - E(\phi(X_n) \mid X_{n-k}) & n > k \\ 0 & n \leq k \end{cases} \end{aligned} \tag{34}$$

Then $\frac{1}{N}\sum_{n=1}^{N} Z_n^k \to 0$ a.s. ν_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: $E\left(Z_n^k \mid Z_{n-1}^k, ..., Z_1^k\right) = 0$. Note that $Z_{n-1}^k, ..., Z_1^k$ are random variables in the σ-algebra of $X_{n-k}, ..., X_1$, so:

$$E\left(Z_n^k \mid Z_{n-1}^k, ..., Z_1^k\right) = E\left(E\left(Z_n^k \mid X_{n-k}, ..., X_1\right) \mid Z_{n-1}^k, ..., Z_1^k\right) \tag{35}$$

Notice that by the Markov property:

$$\begin{aligned} &E\left(Z_n^k \mid X_{n-k}, ..., X_1\right) = E\left(Z_n^k \mid X_{n-k}\right) = \\ &\qquad E(E(\phi(X_n) \mid X_{n-k+1}) \mid X_{n-k}) - E(\phi(X_n) \mid X_{n-k}) = \\ &\qquad E(E(\phi(X_n) \mid X_{n-k+1}, X_{n-k}) \mid X_{n-k}) - E(\phi(X_n) \mid X_{n-k}) = \\ &\qquad\qquad E(\phi(X_n) \mid X_{n-k}) - E(\phi(X_n) \mid X_{n-k}) = 0 \end{aligned} \tag{36}$$

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

Proof. (Breiman's Law of Large Numbers)

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

$$\phi(X_n) - E(\phi(X_n) \mid X_{n-k}) = \sum_{i=1}^{k} Z_n^i \tag{37}$$

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

$$\left|\frac{1}{N}\sum_{n=1}^{N}\phi(X_n) - \frac{1}{N}\sum_{n=k+1}^{N} E(\phi(X_n) \mid X_{n-k})\right| \to 0 \tag{38}$$

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

$$\left|\frac{1}{N}\sum_{n=1}^{N}\phi(X_n) - \frac{1}{N}\sum_{n=1}^{N} E(\phi(X_{n+k}) \mid X_n)\right| \to 0 \tag{39}$$

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

$$\left|\frac{1}{N}\sum_{n=1}^{N}\phi(X_n) - \frac{1}{N}\sum_{n=1}^{N}\frac{1}{M}\sum_{k=1}^{M} E(\phi(X_{n+k}) \mid X_n)\right| \to 0 \tag{40}$$

By Proposition 2, for any $\epsilon > 0$, we can choose an M, s.t. for all n:

$$\left|\frac{1}{M}\sum_{k=1}^{M} E(\phi(X_{n+k}) \mid X_n) - \pi(\phi)\right| < \epsilon \tag{41}$$

As we have uniform convergence.

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

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

As ϵ is arbitrary, taking a countable sequence going to 0 will give us the required result, that ν_x-a.e.:

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

□

4 Appendix

4.1 Proof of Lemma 1

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

$$1 = \mu_n\left(X\right) = \langle \mathbb{1}_X, \mu_n \rangle \to \langle \mathbb{1}_X, \mu \rangle = \mu\left(X\right) \tag{44}$$

And so $\mu\left(X\right) = 1$ (and is a positive functional), and therefore $\mu \in \mathrm{Prob}\left(X\right)$. Therefore $\mathrm{Prob}\left(X\right)$ 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

[Bre60] Leo Breiman. "The Strong Law of Large Numbers for a Class of Markov Chains". In: *Ann. Math. Statist.* 31.3 (Sept. 1960), pp. 801–803. DOI: 10.1214/aoms/1177705810. URL: https://doi.org/10.1214/aoms/1177705810.

[EW11] Manfred Einsiedler and Thomas Ward. *Ergodic theory.* Springer, 2011. Chap. 4.3, pp. 107–108.

[BQ16a] Yves Benoist and Jean-François Quint. *Random Walks on Reductive Groups.* Springer, 2016. Chap. I.3.2, pp. 39–41.

[BQ16b] Yves Benoist and Jean-François Quint. *Random Walks on Reductive Groups.* Springer, 2016. Chap. A.3, pp. 289–293.