STA 711: Probability & Measure Theory
Robert L. Wolpert

8 The Laws of Large Numbers

The traditional interpretation of the *probability* of an event E is its *asymptotic frequency*: the limit as $n \to \infty$ of the fraction of n repeated, similar, and independent trials in which E occurs. Similarly the "expectation" of a random variable X is taken to be its *asymptotic average*, the limit as $n \to \infty$ of the average of n repeated, similar, and independent replications of X. For statisticians trying to make inference about the underlying probability distribution $f(x \mid \theta)$ governing observed random variables X_i, this suggests that we should be interested in the probability distribution for large n of quantities like the sample average of the RVs, $\bar{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$ or the partial sum $S_n := \sum_{i=1}^n X_i$. Three of the most celebrated theorems of probability theory concern this quantity. For independent random variables X_i, all with the same probability distribution satisfying $\mathsf{E}|X_i|^3 < \infty$, set $\mu := \mathsf{E}X_i$, $\sigma^2 := \mathsf{E}|X_i - \mu|^2$, and $S_n := \sum_{i=1}^n X_i$. The three main results are:

Laws of Large Numbers (LLN):

$$\frac{S_n - n\mu}{\sigma n} \longrightarrow 0 \qquad (\textit{pr.} \text{ and } \textit{a.s.})$$

Central Limit Theorem (CLT):

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \Longrightarrow \mathsf{No}(0,1) \qquad (\text{in dist.})$$

Law of the Iterated Logarithm (LIL):

$$\limsup_{n\to\infty} \pm \frac{S_n - n\mu}{\sigma\sqrt{2n\log\log n}} = 1.0 \qquad (\textit{a.s.})$$

Together these three give a clear picture of how quickly and in what sense $\frac{1}{n}S_n$ tends to μ. We begin with the Law of Large Numbers (LLN), first in its "weak" form (asserting convergence *pr.*) and then in its "strong" form (convergence *a.s.*). There are several versions of both theorems. The simplest requires the X_i to be IID and L_2; stronger results allow us to weaken (but not eliminate) the independence requirement, permit non-identical distributions, and consider what happens if we relax the L_2 requirement and allow the RVs to be only L_1 (or worse!).

The text covers these things well; to complement it I am going to: (1) Prove the simplest version, and with it the Borel-Cantelli theorems; and (2) Show what happens with Cauchy random variables, which don't satisfy the requirements (the LLN fails).

8.1 Proofs of the Weak and Strong Laws

Here are two simple versions (one Weak, one Strong) of the Law of Large Numbers; first we prove an elementary but very useful result:

Proposition 1 (Markov's Inequality) *Let $\phi(x) \ge 0$ be non-decreasing on $\mathbb{R}_+$. For any random variable $X \ge 0$ and constant $a \in \mathbb{R}_+$,*

$$\mathsf{P}[X \ge a] \le \mathsf{P}[\phi(X) \ge \phi(a)] \le \mathsf{E}[\phi(X)]/\phi(a)$$

To see this, set $Y := \phi(a)\mathbf{1}_A$ for the event $A := \{\phi(X) \ge \phi(a)\}$ and note $Y \le \phi(X)$ so $\mathsf{E}Y \le \mathsf{E}\phi(X)$.

Theorem 1 (L_2 WLLN) *Let $\{X_n\}$ be independent random variables with the same mean $\mu = \mathsf{E}[X_n]$ and uniformly bounded variance $\mathsf{E}(X_n-\mu)^2 \le B$ for some fixed bound $B < \infty$. Set $S_n := \sum_{j \le n} X_j$ and $\bar X_n := S_n/n = \frac{1}{n}\sum_{j\le n} X_j$. Then, as $n \to \infty$,*

$$(\forall \epsilon > 0) \quad \mathsf{P}[|\bar X_n - \mu| > \epsilon] \to 0. \tag{1}$$

Proof.

$$\mathsf{E}(S_n - n\mu)^2 = \sum_{i=1}^n \mathsf{E}(X_i - \mu)^2 \le n\,B$$

so for $\epsilon > 0$

$$\begin{aligned} \mathsf{P}[|\bar X_n - \mu| > \epsilon] &= \mathsf{P}[(S_n - n\mu)^2 > (n\epsilon)^2] \\ &\le \mathsf{E}[(S_n - n\mu)^2]/n^2\epsilon^2 \\ &\le B/n\,\epsilon^2 \to 0 \quad \text{as } n \to \infty. \end{aligned}$$

∎

This Law of Large Numbers is called *weak* because its conclusion is only that $\bar X_n$ converges to zero *in probability* (Eqn (1)); the *strong* Law of Large Numbers asserts convergence of a stronger sort, called *almost sure* convergence (Eqn (2) below). If $\mathsf{P}[|\bar X_n - \mu| > \epsilon]$ were *summable* then by B-C we could conclude almost-sure convergence; unfortunately we have only the bound $\mathsf{P}[|\bar X_n - \mu| > \epsilon] < c/n$ which tends to zero but isn't summable. It is summable along the *subsequence* n^2, however; our approach to proving a strong LLN is to show that $|S_k - S_{n^2}|$ isn't too big for any $n^2 \le k < (n+1)^2$.

Theorem 2 (L_2 SLLN) *Under the same conditions,*

$$\mathsf{P}[\bar X_n \to \mu] = 1. \tag{2}$$

Proof. Without loss of generality take $\mu = 0$ (otherwise subtract μ from each X_n), and fix $\epsilon > 0$. Set $S_n := \sum_{j\le n} X_j$. Then

$$\begin{aligned} \mathsf{P}\big[|S_{n^2}| > n^2\epsilon\big] &\le \mathsf{E}|S_{n^2}|^2/(n^2\epsilon)^2 \le n^2 B/n^4\epsilon^2 = B/n^2\epsilon^2 \\ \mathsf{P}\big[|S_{n^2}| > n^2\epsilon \ \text{ i.o.}\big] &= 0 \text{ by B-C } \Rightarrow S_{n^2}/n^2 \to 0 \text{ a.s. Set} \\ D_n &:= \max_{n^2 \le k < (n+1)^2} |S_k - S_{n^2}| \\ \mathsf{E}D_n^2 &= \mathsf{E}\left[\max_{n^2 \le k < (n+1)^2} |S_k - S_{n^2}|^2\right] \\ &\le \mathsf{E}\sum_{n^2 \le k < (n+1)^2} |S_k - S_{n^2}|^2 = \sum_{n^2 \le k < (n+1)^2} \mathsf{E}|S_k - S_{n^2}|^2 \\ &\le \sum_{n^2 \le k < (n+1)^2} (k - n^2)B \le 4n^2 B\ , \text{ so} \\ \mathsf{P}[D_n > n^2\epsilon] &\le 4n^2 B/n^4\,\epsilon^2 = 4B/n^2\,\epsilon^2 \Rightarrow D_n/n^2 \to 0 \text{ a.s.} \\ \left|\frac{S_k}{k}\right| &\le \frac{|S_{n^2}| + D_n}{n^2} \to 0 \text{ a.s. as } k \to \infty, \text{ where } n = \lfloor \sqrt{k} \rfloor. \end{aligned}$$

$\square$

Each of these LLNs required only that $\mathsf{Cov}(X_n, X_m) \le 0$, not pairwise (let alone full) independence. To see that, suppose $\mathsf{E}X_n = \mu$ for all n, $\mathsf{Cov}(X_n, X_m) \le 0$ for $n \ne m$, and $\mathsf{V}X_n \le B$. Then

$$\begin{aligned}
\mathsf{E}(S_n - n\mu)^2 &= \mathsf{E} \sum_{i,j=1,1}^{n,n} (X_i - \mu)(X_j - \mu) \\
&= \sum_{i=1}^{n} \mathsf{E}(X_i - \mu)^2 + 2 \sum_{1 \le i < j \le n} \mathsf{E}(X_i - \mu)(X_j - \mu) \\
&\le nB.
\end{aligned}$$

We'll see below in Section (8.4) that even positive correlations are okay if they fall off fast enough— *e.g.*, if $|\mathsf{Cov}(X_n, X_m)| \le ar^{|n-m|}$ for some $a > 0$, $0 < r < 1$— with a similar proof. The uniform L_2 bound isn't necessary either. There are a variety of LLNs with either or both of the L_2 bound and independence weakened in some way, but they can't be dispensed with altogether— consider iid Cauchy random variables, for example, to show L_2 isn't entirely superfluous, or $X_n \equiv X_1$ with any nontrivial distribution to show the need for at least a modicum of independence.

8.2 Other Strong Laws

Let's first state two lemmas:

Lemma 1 (Lévy) *If $\{X_n\}$ is an independent sequence then $\sum_{n=1}^{\infty} X_n$ converges (*pr.*) if and only if it converges* a.s.

Lemma 2 (Kronecker) *Suppose $\{x_n\} \subset \mathbb{R}$ and $0 < a_n \nearrow \infty$. Then*

$$\sum_{k=1}^{\infty} \frac{x_k}{a_k} \text{ converges } \Rightarrow \frac{1}{a_n} \sum_{k=1}^{n} x_k \to 0.$$

and, from these, prove two useful theorems. First,

Theorem 3 (Kolmogorov Convergence Criterion) *Let $\{X_n\} \subset L_2$ be an independent sequence with means $\mu_n := \mathsf{E}X_n$ and variances $\sigma_n^2 := \mathsf{V}(X_n)$. Then*

$$\sum_{n=1}^{\infty} \sigma_n^2 < \infty \;\Rightarrow\; \sum_{n=1}^{\infty} (X_n - \mu_n) \text{ converges a.s.}$$

Proof. Without loss of generality take $\mu_n \equiv 0$. Fix $\epsilon > 0$. For $1 \le n \le N < \infty$,

$$\begin{aligned}
\mathsf{P}\big[\big|S_N - S_n\big| > \epsilon\big] &= \mathsf{P}\big[\big|\sum_{n<k\le N} X_k\big| > \epsilon\big] \\
&\le \mathsf{E}\big(\sum_{n<k\le N} X_k\big)^2/\epsilon^2 && \text{by Markov's inequality} \\
&\le \sum_{n<k\le\infty} \sigma_k^2/\epsilon^2 && \text{by independence} \\
&\to 0 \quad \text{as } n \to \infty && \text{since } \sum \sigma_n^2 < \infty,
\end{aligned}$$

so S_n is a convergent Cauchy sequence in probability and, by Lemma 1, also converges *a.s.* ☐

Theorem 4 (Kronecker L_2 SLLN) *Let $\{X_n\} \subset L_2$ be an independent sequence, and let $\{b_n\} \subset \mathbb{R}_+$ be a monotone sequence such that*

$$\sum_n \mathsf{V}\Big(\frac{X_n}{b_n}\Big) < \infty.$$

Then

$$\frac{S_n - \mathsf{E}S_n}{b_n} \to 0 \text{ a.s.}$$

In particular, for iid $\{X_n\}$ we may take $b_n = n$ to see $\bar{X}_n \to \mu$ *a.s.*, a SLLN for independent but non-identically-distributed L_2 random variables.

Proof. Again take $\mathsf{E}[X_n] \equiv 0$, and set $Y_n := X_n/b_n$, with variance σ_n^2. By hypothesis $\sum \sigma_n^2 < \infty$, so $\sum Y_n$ coverges *a.s.* by Theorem 3. Thus, for almost every ω, $\sum X_n(\omega)/b_n$ converges and, by Lemma 2, also $S_n/b_n \to 0$. ☐

Kolmogorov's L_1 SLLN

The most-cited and most-used version of the Strong Law for iid sequences is that due to Kolmogorov, with no moment assumptions:

Theorem 5 (Kolmogorov's L_1 SLLN) *Let $\{X_n\}$ be iid and set $S_n := \sum_{i\le n} X_i$. There exists $c \in \mathbb{R}$ such that*

$$\bar{X}_n = S_n/n \to c \quad \text{a.s.}$$

if and only if $\mathsf{E}|X_1| < \infty$, in which case $c = \mathsf{E}X_1$.

This has the pleasant consequence that the usual estimators for the mean and variance of iid sequences are consistent:

Corollary 1

$$\{X_n\} \subset L_1 \Rightarrow \bar{X}_n \to \mu \quad \text{a.s.}$$
$$\{X_n\} \subset L_2 \Rightarrow \frac{1}{n}\sum_{i\le n}(X_n - \bar{X}_n)^2 \to \sigma^2 \quad \text{a.s.}$$

Here's a quick summary of some LLN facts:

I. Weak version, non-*iid*, L_2: $\mu_i = \mathsf{E}X_i$, $\sigma_{ij} = \mathsf{E}[X_i - \mu_i][X_j - \mu_j]$

  A. $Y_n = (S_n - \Sigma\mu_i)/n$ satisfies $\mathsf{E}Y_n = 0$, $\mathsf{E}Y_n^2 = \frac{1}{n^2}\Sigma_{i\le n}\sigma_{ii} + \frac{2}{n^2}\Sigma_{i<j\le n}\sigma_{ij}$;

    1. If $\sigma_{ii} \le M$ and $\sigma_{ij} \le 0$ *or* $|\sigma_{ij}| < Mr^{|i-j|}$, $r < 1$,
    Chebychev $\Longrightarrow Y_n \to 0$ $(pr.)$

    2. (pairwise) IID L_2 is OK

II. Strong version, non-*iid*, L_2: $\mathsf{E}X_i = 0$, $\mathsf{E}X_i^2 \le M$, $\mathsf{E}X_iX_j \le 0$.

  A. $\mathsf{P}[|S_n| > n\epsilon] < \frac{Mn}{n^2\epsilon^2} = \frac{M}{n\epsilon^2}$

    1. $\mathsf{P}[|S_{n^2}| > n^2\epsilon] < \frac{M}{n^2\epsilon^2}$, $\Sigma_n \mathsf{P}[|S_{n^2}| > n^2\epsilon] < \frac{M\pi^2}{6\epsilon^2} < \infty$

    2. Borel-Cantelli: $\mathsf{P}[|S_{n^2}| > n^2\epsilon\ i.o.] = 0$, $\therefore \frac{1}{n^2}S_{n^2} \to 0$ $a.s.$

    3. $D_n := \max_{n^2 \le k < (n+1)^2} |S_k - S_{n^2}|$, so
    $\mathsf{E}D_n^2 \le 2n\mathsf{E}|S_{(n+1)^2-1} - S_{n^2}| \le 4n^2M$

    4. Chebychev: $\mathsf{P}[D_n > n^2\epsilon] < \frac{4n^2M}{n^4\epsilon^2}$, $\therefore D_n/n^2 \to 0$ $a.s.$

  B. $|S_k/k| \le \frac{|S_{n^2}| + D_n}{n^2} \to 0$ $a.s.$ as $k \to \infty$, QED

    1. Bernoulli RVs, normal number theorem, Monte Carlo integration.

III. Weak version, pairwise-*iid*, L_1

  A. Equivalent sequences: If $\sum_n \mathsf{P}[X_n \ne Y_n] < \infty$, then:

    1. $\sum_n |X_n - Y_n| < \infty$ $a.s.$

    2. $\sum_{i=1}^n X_i$ converges iff $\sum_{i=1}^n Y_i$ converges

    3. If $(\exists a_n \nearrow)(\exists X) \ni \frac{1}{a_n}\sum_{i\le n} X_i \to X$ then $\frac{1}{a_n}\sum_{i\le n} Y_i \to X$ too.

  B. If $\{X_n\}$ are iid L_1 then X_n, $Y_n := X_n 1_{[|X_n| \le n]}$ are equivalent

IV. Strong version, *iid*, L_1

  A. Kolmogorov: For $\{X_n\}$ IID, $(\exists c \ni \bar{X}_n \to c\ a.s.) \Leftrightarrow (X_n \in L_1)$.

V. Strong version, non-*iid*, L_p

  A. Let $\{X_n\}$ be independent and uniformly bounded in L_p for some $p > 1$. Then $(1/n)\sum_{j=1}^n (X_j - \mu_j) \to 0$ $a.s.$

VI. Counterexamples: Cauchy,

  A. $X_i \sim \frac{dx}{\pi[1+x^2]} \Longrightarrow \mathsf{P}[|S_n|/n \le \epsilon] \equiv \frac{2}{\pi}\tan^{-1}(\epsilon) \not\to 1$, WLLN fails.

  B. $\mathsf{P}[X_i = \pm n] = \frac{c}{n^2}$, $n \ge 1$; $X_i \notin L_1$, and $S_n/n \not\to 0$ $(pr.)$ or $a.s.$

  C. $\mathsf{P}[X_i = \pm n] = \frac{c}{n^2 \log n}$, $n > 1$; $X_i \notin L_1$, but $S_n/n \to 0$ $(pr.)$ and not $a.s.$

  D. Medians: for ANY RVs $X_n \to X_\infty$ $pr.$, then $m_n \to m_\infty$ if m_∞ is unique.

8.3 An example where the LLN fails: iid Cauchy RVs

Let X_i be iid standard Cauchy RVs, with

$$\mathsf{P}[X_1 \le t] = \int_{-\infty}^{t} \frac{dx}{\pi[1+x^2]} = \frac{1}{2} + \frac{1}{\pi}\arctan(t)$$

and characteristic function (we'll learn more about these next week)

$$\mathsf{E}\, e^{i\omega X_1} = \int_{-\infty}^{\infty} e^{i\omega x} \frac{dx}{\pi[1+x^2]} = e^{-|\omega|}.$$

The sample mean $\bar{X}_n := S_n/n$ has characteristic function

$$\mathsf{E}\, e^{i\omega S_n/n} = \mathsf{E}\, e^{i\frac{\omega}{n}[X_1+\cdots+X_n]} = \left(\mathsf{E}\, e^{i\frac{\omega}{n}X_1}\right)^n = (e^{-|\frac{\omega}{n}|})^n = e^{-|\omega|}.$$

Thus S_n/n also has the standard Cauchy distribution with $\mathsf{P}[S_n/n \le t] = \frac{1}{2} + \frac{1}{\pi}\arctan(t)$. In particular, S_n/n does not converge almost surely, or even in probability.

8.4 A LLN for Correlated Sequences

In many applications we would like a Law of Large Numbers for sequences of random variables that are *not* independent. For example, in Markov Chain Monte Carlo integration, we have a stationary Markov chain $\{X_t\}$ (this means that the distribution of X_t is the same for all t and that the conditional distribution of a future value X_u for $u > t$, given the past $\{X_s \mid s \le t\}$, depends only on the present X_t) and want to estimate the population mean $\mathsf{E}[\phi(X_t)]$ for some function $\phi(\cdot)$ by the sample mean

$$\mathsf{E}[\phi(X_t)] \approx \frac{1}{T}\sum_{t=0}^{T-1} \phi(X_t).$$

Even though they are identically distributed, the random variables $Y_t := \phi(X_t)$ won't be independent if the X_t aren't independent, so the LLN we already have doesn't quite apply.

Theorem 6 *If an L_2 sequence $\{Y_t\}$ has a constant mean $\mu = \mathsf{E}Y_t$ and a summable autocovariance $\gamma_{st} := \mathsf{E}(Y_s-\mu)(Y_t-\mu)$ that satisfies $\sum_{t=-\infty}^{\infty} |\gamma_{st}| \le c < \infty$ uniformly in s, then the random variables $\{Y_t\}$ obey a LLN:*

$$\mathsf{E}[Y_t] = \lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} Y_t.$$

Proof. Let S_T be the sum of the first T Y_ts and (as usual) set $\bar{Y}_T := S_T/T$. The variance of S_T is

$$\begin{aligned}\mathsf{E}[(S_T - T\mu)^2] &= \sum_{s=0}^{T-1}\sum_{t=0}^{T-1} \mathsf{E}[(Y_s-\mu)(Y_t-\mu)]\\ &\le \sum_{s=0}^{T-1}\sum_{t=-\infty}^{\infty} |\gamma_{st}| \le T\,c,\end{aligned}$$

so $\bar{Y}_T$ has variance $\mathsf{V}[\bar{Y}_T] \le c/T$ and by Chebychev's inequality

$$\begin{aligned}\mathsf{P}[|\bar{Y}_T - \mu| > \epsilon] &\le \frac{\mathsf{E}[(\bar{Y}_T - \mu)^2]}{\epsilon^2} = \frac{\mathsf{E}[(S_T - T\mu)^2]}{T^2\epsilon^2} \\ &\le \frac{T\,c}{T^2\epsilon^2} = \frac{c}{T\epsilon^2} \to 0 \quad \text{as } T \to \infty.\end{aligned}$$

As in the iid case, this WLLN can be extended to a SLLN by first applying it along the sequence $\{T^2 :\ T \in \mathbb{N}\}$, applying B/C, then filling in the gaps.

A sequence of random variables Y_t is called *stationary* if each Y_t has the same probability distribution and, moreover, each finite set $(Y_{t_1+h}, Y_{t_2+h}, \cdots, Y_{t_k+h})$ has a joint distribution that doesn't depend on h. The sequence is called "L_2" if each Y_t has a finite variance σ^2 (and hence also a well-defined mean μ); by stationarity it also follows that the *covariance*

$$\gamma_{st} = \mathsf{E}[(Y_s - \mu)(Y_t - \mu)]$$

is finite and depends only on the absolute difference $|t - s|$ (write: $\gamma_{st} = \gamma(s - t) = \gamma(t - s)$).

Corollary 2 *If a stationary L_2 sequence $\{Y_t\}$ has a summable covariance function,* i.e.*, satisfies $\sum_{t=-\infty}^{\infty} |\gamma(t)| \le c < \infty$, then*

$$\mathsf{E}[Y_t] = \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} Y_t.$$

Note we didn't need full stationarity. It would be good enough for $\{Y_t\}$ to be "2^{nd}-order stationary," *i.e.*, to have a common mean $\mu = \mathsf{E}Y_t$ and a covariance function $\gamma_{st} = \mathsf{E}(Y_s - \mu)(Y_t - \mu) = \gamma(s - t)$ that depends only on the absolute difference $|s - t|$.

8.4.1 Examples

1. **IID:** If X_t are independent and identically distributed, and if $Y_t = \phi(X_t)$ has finite variance σ^2, then Y_t has a well-defined finite mean μ and $\bar{Y}_T \to \mu$.
   Here $\gamma_{st} = \begin{cases}\sigma^2 & \text{if } s = t \\ 0 & \text{if } s \ne t\end{cases}$, so $c = \sigma^2 < \infty$.

2. **AR$_1$:** If Z_t are *iid* $\mathsf{No}(0, 1)$ for $-\infty < t < \infty$, $\mu \in \mathbb{R}$, $\sigma > 0$, $-1 < \rho < 1$, and

$$\begin{aligned}X_t &:= \mu + \sigma \sum_{s=0}^{\infty} \rho^s Z_{t-s} \\ &= \rho X_{t-1} + \alpha + \sigma Z_t, \qquad (*)\end{aligned}$$

   where $\alpha = (1 - \rho)\mu$, then the X_t are identically distributed (all with the $\mathsf{No}(\mu, \frac{\sigma^2}{1-\rho^2})$ distribution) but not independent (since $\gamma_{st} = \frac{\sigma^2}{1-\rho^2}\rho^{|s-t|} \ne 0$ for $s \ne t$). This is called an "autoregressive process" (because of equation (*), expressing X_t as a regression of previous

X_ss) of order one (because only one earlier X_s appears in (*)), and is about the simplest non-*iid* sequence occuring in applications. Since the covariance is summable,

$$\sum_{t=-\infty}^{\infty} |\gamma(t)| = \frac{\sigma^2}{1-\rho^2}\frac{1+|\rho|}{1-|\rho|} = \frac{\sigma^2}{(1-|\rho|)^2} < \infty,$$

we again have $\bar{X}_T \to \mu$ *a.s.* as $T \to \infty$.

3. **Geometric Ergodicity:** If for some $0 < \rho < 1$ and $c > 0$ we have $\gamma_{st} \le c\rho^{|s-t|}$ for a Markov chain Y_t the chain is called *Geometrically Ergodic* (because $c\rho^t$ is a geometric sequence), and the same argument as for AR_1 shows that $\bar{Y}_t$ converges. Meyn & Tweedie (1993), Tierney (1994), and others have given conditions for MCMC chains to be Geometric Ergodic, and hence for the almost-sure convergence of sample averages to population means. The mean squared error $\mathrm{MSE} := \mathsf{E}|\bar{Y}_t - \mu|^2$ for a geometrically ergodic sequence is bounded by $\mathrm{MSE} \le \frac{1+\rho}{1-\rho}(\sigma^2/t) \asymp 1/t$, but the constant grows without bound as $\rho \to 1$. Irreducible aperiodic finite-state Markov chains are geometrically ergodic, with ρ the second-largest eigenvalue of the one-step transition matrix.

4. **General Ergodicity:** Consider the three sequences of random variables on $(\Omega, \mathcal{F}, \mathsf{P})$ with $\Omega = (0, 1]$ and $\mathcal{F} = \mathcal{B}(\Omega)$, each with $X_0(\omega) = \omega$:

   (a) $X_{n+1} := 2\,X_n \pmod 1$;

   (b) $X_{n+1} := X_n + \alpha \pmod 1$ (Does it matter if α is rational?);

   (c) $X_{n+1} := 4X_n(1 - X_n)$.

   For each there exists a probability measure P (a distribution for X_0) such that the X_n are identically distributed— the $\mathsf{Un}(0, 1)$ distribution for (a) and (b), and the "arcsin" distribution $\mathsf{Be}(\frac{1}{2}, \frac{1}{2})$ for (c), and each is of the form $X_n = X_0\big(T^n(\omega)\big)$ for a measure-preserving transformation $T : \Omega \to \Omega$, a measurable mapping $T : \Omega \to \Omega$ for which $\mathsf{P}(E) = \mathsf{P}\big(T^{-1}(E)\big)$ for each $E \in \mathcal{F}$. Such a transformation is called *ergodic* if the only T-invariant events (those that satisfy $E = T^{-1}(E)$) are "almost trivial" in the sense that $\mathsf{P}[E] = 0$ or $\mathsf{P}[E] = 1$. A sequence $X_n := X_0 \circ T^n$ is called ergodic if T is. Both sequences (a) and (c) are ergodic (can you show that?); sequence (b) is ergodic if and only if α is irrational (what happens if $\alpha \in \mathbb{Q}$?)

   Birkhoff's *Ergodic Theorem* asserts that, if $X_n \in L_1(\Omega, \mathcal{F}, \mathsf{P})$ is an integrable ergodic sequence, then $\bar{X}_n$ converges almost-surely to a T-invariant limit X_∞ as $n \to \infty$. Since only constants are T-invariant for ergodic sequences, it follows that $\bar{X}_n \to \mu := \mathsf{E}X_n$ *a.s.* for ergodic sequences. The conditions here are weaker than those for the usual LLN; in all three cases above, for example, all ergodic, each X_n is completely determined by X_0 so there is complete dependence, with $\sigma(X_n) \subset \sigma(X_m) \subset \sigma(X_0)$ for all $0 \le m \le n$!

   For any L_1 distribution $\mu(dx)$ on $(\mathbb{R}, \mathcal{B})$, we can construct iid random variables $\{X_n\} \overset{\text{iid}}{\sim} \mu(dx)$ on the product probability space $\big(\Omega = \mathbb{R}^\infty, \mathcal{F} = \mathcal{B}^\infty, \mathsf{P} = \bigotimes \mu\big)$ and a measure-preserving transformation $T : \Omega \to \Omega$ called the *left-shift* by

$$X_n(\omega_1, \omega_2, \omega_3, \cdots) := \omega_n \qquad\qquad T(\omega_1, \omega_2, \omega_3, \cdots) := (\omega_2, \omega_3, \omega_4, \cdots).$$

The σ-algebra $\mathcal{T} = \{A: \ A = T^{-1}(A)\}$ of T-invariant sets is just the tail σ-algebra for the independent random variables $\{X_n\}$, so by Kolmogorov's zero-one law $\mathcal{T}$ is almost-trivial and so T is ergodic. It follows from Birkhoff's Ergodic Theorem that the sample mean $\bar{X}_n := \frac{1}{n}\sum_{j=1}^n X_j$ converges almost-surely to a T-invariant and hence almost-surely constant random variable whose value must be μ, proving a strong LLN for iid random variables that assumes only L_1:

Theorem 7 (L_1 iid SLLN) *Let $\{X_n\}$ be iid $L_1(\Omega, \mathcal{F}, \mathsf{P})$ random variables with mean $\mu = \mathsf{E}[X_n]$. Set $S_n := \sum_{j\le n} X_j$ and $\bar{X}_n := S_n/n = \frac{1}{n}\sum_{j\le n} X_j$. Then:*

$$\mathsf{P}[\bar{X}_n \to \mu] = 1.$$

Thus the "space average" $\int_{\mathbb{R}} X_n \, d\mu$ and the limiting "time average" $\lim_{n\to\infty} \bar{X}_n$ coincide for ergodic sequences.

Last edited: October 23, 2019