Lecture 9

The Strong Law of Large Numbers

Reading: Grimmett-Stirzaker 7.2; David Williams "Probability with Martingales" 7.2
Further reading: Grimmett-Stirzaker 7.1, 7.3-7.5

With the Convergence Theorem (Theorem 54) and the Ergodic Theorem (Theorem 55) we have two very different statements of convergence of something to a stationary distribution. We are looking at a recurrent Markov chain $(X_t)_{t\geq 0}$, i.e. one that visits *every* state at arbitrarily large times, so clearly X_t itself does not converge, as $t \to \infty$. In this lecture, we look more closely at the different types of convergence and develop methods to show the so-called almost sure convergence, of which the statement of the Ergodic Theorem is an example.

9.1 Modes of convergence

Definition 59 Let X_n, $n \geq 1$, and X be random variables. Then we define

1. $X_n \to X$ *in probability*, if for all $\varepsilon > 0$, $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$ as $n \to \infty$.
2. $X_n \to X$ *in distribution*, if $\mathbb{P}(X_n \leq x) \to \mathbb{P}(X \leq x)$ as $n \to \infty$, for all $x \in \mathbb{R}$ at which $x \mapsto \mathbb{P}(X \leq x)$ is continuous.
3. $X_n \to X$ *in* L^1, if $\mathbb{E}(|X_n|) < \infty$ for all $n \geq 1$ and $\mathbb{E}(|X_n - X|) \to 0$ as $n \to \infty$.
4. $X_n \to X$ *almost surely* (a.s.), if $\mathbb{P}(X_n \to X \text{ as } n \to \infty) = 1$.

Almost sure convergence is the notion that we will study in more detail here. It helps to consider random variables as functions $X_n : \Omega \to \mathbb{R}$ on a sample space Ω, or at least as functions of a common, typically infinite, family of independent random variables. What is different here from previous parts of the course (except for the Ergodic Theorem, which we yet have to inspect more thoroughly), is that we want to calculate probabilities that fundamentally depend on an infinite number of random variables. So far, we have been able to revert to events depending on only finitely many random variables by conditioning. This will not work here.

Let us start by recalling the definition of convergence of sequences, as $n \to \infty$,

$$x_n \to x \qquad \iff \qquad \forall_{m\geq 1}\exists_{n_m\geq 1}\forall_{n\geq n_m}|x_n - x| < 1/m.$$

If we want to consider all sequences $(x_n)_{n\geq 1}$ of possible values of the random variables $(X_n)_{n\geq 1}$, then

$$n_m = \inf\{k \geq 1 : \forall_{n\geq k}|x_n - x| < 1/m\} \in \mathbb{N} \cup \{\infty\}$$

will vary as a function of the sequence $(x_n)_{n\geq 1}$, and so it will become a random variable

$$N_m = \inf\{k \geq 1 : \forall_{n\geq k}|X_n - X| < 1/m\} \in \mathbb{N} \cup \{\infty\}$$

as a function of $(X_n)_{n\geq 1}$. This definition of N_m permits us to write

$$\mathbb{P}(X_n \to X) = \mathbb{P}(\forall_{m\geq 1}N_m < \infty).$$

This will occasionally help, when we are given almost sure convergence, but is not much use when we want to prove almost sure convergence. To prove almost sure convergence, we can transform as follows

$$\begin{aligned} &\mathbb{P}(X_n \to X) = \mathbb{P}(\forall_{m\geq 1}\exists_{N\geq 1}\forall_{n\geq N}|X_n - X| < 1/m) = 1 \\ \iff \quad &\mathbb{P}(\exists_{m\geq 1}\forall_{N\geq 1}\exists_{n\geq N}|X_n - X| \geq 1/m) = 0. \end{aligned}$$

We are used to events such as $A_{m,n} = \{|X_n - X| \geq 1/m\}$, and we understand events as subsets of Ω, or loosely identify this event as set of all $((x_k)_{k\geq 1}, x)$ for which $|x_n - x| \geq 1/m$. This is useful, because we can now translate $\exists_{m\geq 1}\forall_{N\geq 1}\exists n \geq N$ into set operations and write

$$\mathbb{P}(\cup_{m\geq 1} \cap_{N\geq 1} \cup_{n\geq N} A_{m,n}) = 0.$$

This event can only have zero probability if all events $\cap_{N\geq 1} \cup_{n\geq N} A_{m,n}$, $m \geq 1$, have zero probability (formally, this follows from the sigma-additivity of the measure $\mathbb{P}$). The Borel-Cantelli lemma will give a criterion for this.

Proposition 60 *The following implications hold*

$$\begin{array}{ccccc} X_n \to X \ \textit{almost surely} & & \\ \Downarrow & & \\ X_n \to X \ \textit{in probability} & \Rightarrow & X_n \to X \ \textit{in distribution} \\ \Uparrow & & \\ X_n \to X \ \textit{in } L^1 & \Rightarrow & \mathbb{E}(X_n) \to \mathbb{E}(X) \end{array}$$

No other implications hold in general.

Proof: Most of this is Part A material. Some counterexamples are on Assignment 5. It remains to prove that almost sure convergence implies convergence in probability. Suppose, $X_n \to X$ almost surely, then the above considerations yield $\mathbb{P}(\forall_{m\geq 1}N_m < \infty) = 1$, i.e. $\mathbb{P}(N_k < \infty) \geq \mathbb{P}(\forall_{m\geq 1}N_m < \infty) = 1$ for all $k \geq 1$.

Now fix $\varepsilon > 0$. Choose $m \geq 1$ such that $1/m < \varepsilon$. Then clearly $|X_n - X| > \varepsilon > 1/m$ implies $N_m > n$ so that

$$\mathbb{P}(|X_n - X| > \varepsilon) \leq \mathbb{P}(N_m > n) \to \mathbb{P}(N_m = \infty) = 0,$$

as $n \to \infty$, for any $\varepsilon > 0$. Therefore, $X_n \to X$ in probability. □

9.2 The first Borel-Cantelli lemma

Let us now work on a sample space Ω. It is safe to think of $\Omega = \mathbb{R}^{\mathbb{N}} \times \mathbb{R}$ and $\omega \in \Omega$ as $\omega = ((x_n)_{n\geq 1}, x)$ as the set of possible outcomes for an infinite family of random variables (and a limiting variable).

The Borel-Cantelli lemmas are useful to prove almost sure results. Particularly limiting results often require certain events to happen infinitely often (i.o.) or only a finite number of times. Logically, this can be expressed as follows. Consider events $A_n \subset \Omega$, $n \geq 1$. Then

$$\omega \in A_n \text{ i.o.} \quad \Longleftrightarrow \quad \forall_{n\geq 1} \exists_{m\geq n} \quad \omega \in A_m \quad \Longleftrightarrow \quad \omega \in \bigcap_{n\geq 1} \bigcup_{m\geq n} A_m$$

Lemma 61 (Borel-Cantelli (first lemma)) *Let $A = \bigcap_{n\geq 1} \bigcup_{m\geq n} A_m$ be the event that infinitely many of the events A_n occur. Then*

$$\sum_{n\geq 1} \mathbb{P}(A_n) < \infty \quad \Rightarrow \quad \mathbb{P}(A) = 0$$

Proof: We have that $A \subset \bigcup_{m\geq n} A_m$ for all $n \geq 1$, and so

$$\mathbb{P}(A) \leq \mathbb{P}\left(\bigcup_{m\geq n} A_m \right) \leq \sum_{m\geq n} \mathbb{P}(A_m) \to 0 \qquad \text{as } n \to \infty$$

whenever $\sum_{n\geq 1} \mathbb{P}(A_n) < \infty$. □

9.3 The Strong Law of Large Numbers

Theorem 62 *Let $(X_n)_{n\geq 1}$ be a sequence of independent and identically distributed (iid) random variables with $\mathbb{E}(X_1^4) < \infty$ and $\mathbb{E}(X_1) = \mu$. Then*

$$\frac{S_n}{n} := \frac{1}{n} \sum_{i=1}^{n} X_i \to \mu \qquad \textit{almost surely.}$$

Fact 63 *Theorem 62 remains valid without the assumption $\mathbb{E}(X_1^4) < \infty$, just assuming $\mathbb{E}(|X_1|) < \infty$.*

The proof for the general result is hard, but under the extra moment condition $\mathbb{E}(X_1^4) < \infty$ there is a nice proof.

Lemma 64 *In the situation of Theorem 62, there is a constant $K < \infty$ such that for all $n \geq 0$*

$$\mathbb{E}((S_n - n\mu)^4) \leq Kn^2.$$

Proof: Let $Z_k = X_k - \mu$ and $T_n = Z_1 + \ldots + Z_n = S_n - n\mu$. Then

$$\mathbb{E}(T_n^4) = \mathbb{E}\left(\left(\sum_{i=1}^n Z_i\right)^4\right) = n\mathbb{E}(Z_1^4) + 3n(n-1)\mathbb{E}(Z_1^2 Z_2^2) \leq Kn^2$$

by expanding the fourth power and noting that most terms vanish such as

$$\mathbb{E}(Z_1 Z_2^3) = \mathbb{E}(Z_1)\mathbb{E}(Z_2^3) = 0.$$

K was chosen appropriately, say $K = 4\max\{\mathbb{E}(Z_1^4), (\mathbb{E}(Z_1^2))^2\}$. □

Proof of Theorem 62: By the lemma,

$$\mathbb{E}\left(\left(\frac{S_n}{n} - \mu\right)^4\right) \leq Kn^{-2}$$

Now, by Tonelli's theorem,

$$\mathbb{E}\left(\sum_{n\geq 1}\left(\frac{S_n}{n} - \mu\right)^4\right) = \sum_{n\geq 0}\mathbb{E}\left(\left(\frac{S_n}{n} - \mu\right)^4\right) < \infty \quad \Rightarrow \quad \sum_{n\geq 1}\left(\frac{S_n}{n} - \mu\right)^4 < \infty \quad \text{a.s.}$$

But if a series converges, the underlying sequence converges to zero, and so

$$\left(\frac{S_n}{n} - \mu\right)^4 \to 0 \quad \text{almost surely} \qquad \Rightarrow \qquad \frac{S_n}{n} \to \mu \quad \text{almost surely.}$$

□

This proof did not use the Borel-Cantelli lemma, but we can also conclude by the Borel-Cantelli lemma:

Proof of Theorem 62: We know by Markov's inequality that

$$\mathbb{P}\left(\frac{1}{n}|S_n - n\mu| \geq n^{-\gamma}\right) \leq \frac{\mathbb{E}((S_n/n - \mu)^4)}{n^{-4\gamma}} = Kn^{-2+4\gamma}.$$

Define for $\gamma \in (0, 1/4)$

$$A_n = \left\{\frac{1}{n}|S_n - n\mu| \geq n^{-\gamma}\right\} \quad \Rightarrow \quad \sum_{n\geq 1}\mathbb{P}(A_n) < \infty \quad \Rightarrow \quad \mathbb{P}(A) = 0$$

by the first Borel-Cantelli lemma, where $A = \bigcap_{n\geq 1}\bigcup_{m\geq n} A_n$. But now, event A^c happens if and only if

$$\exists_N \forall_{n\geq N}\left|\frac{S_n}{n} - \mu\right| < n^{-\gamma} \qquad \Rightarrow \qquad \frac{S_n}{n} \to \mu.$$

□

9.4 The second Borel-Cantelli lemma

We won't need the second Borel-Cantelli lemma in this course, but include it for completeness.

Lemma 65 (Borel-Cantelli (second lemma)) *Let* $A = \bigcap_{n\geq 1}\bigcup_{m\geq n} A_n$ *be the event that infinitely many of the events* A_n *occur. Then*

$$\sum_{n\geq 1}\mathbb{P}(A_n) = \infty \text{ and } (A_n)_{n\geq 1} \text{ independent} \quad\Rightarrow\quad \mathbb{P}(A) = 1.$$

Proof: The conclusion is equivalent to $\mathbb{P}(A^c) = 0$. By de Morgan's laws

$$A^c = \bigcup_{n\geq 1}\bigcap_{m\geq n} A_m^c.$$

However,

$$\begin{aligned}\mathbb{P}\left(\bigcap_{m\geq n} A_m^c\right) &= \lim_{r\to\infty}\mathbb{P}\left(\bigcap_{m=n}^{r} A_m^c\right)\\ &= \prod_{m\geq n}(1-\mathbb{P}(A_m)) \leq \prod_{m\geq n}\exp\left(-\mathbb{P}(A_m)\right) = \exp\left(-\sum_{m\geq n}\mathbb{P}(A_m)\right) = 0\end{aligned}$$

whenever $\sum_{n\geq 1}\mathbb{P}(A_n) = \infty$. Thus

$$\mathbb{P}(A^c) = \lim_{n\to\infty}\mathbb{P}\left(\bigcap_{m\geq n} A_m^c\right) = 0.$$

□

As a technical detail: to justify some of the limiting probabilities, we use "continuity of $\mathbb{P}$" along increasing and decreasing sequences of events, that follows from the sigma-additivity of $\mathbb{P}$, cf. Grimmett-Stirzaker, Lemma 1.3.(5).

9.5 Examples

Example 66 (Arrival times in Poisson process) A Poisson process has independent and identically distributed inter-arrival times $(Z_n)_{n\geq 0}$ with $Z_n \sim Exp(\lambda)$. We denoted the partial sums (arrival times) by $T_n = Z_0 + \ldots + Z_{n-1}$. The Strong Law of Large Numbers yields

$$\frac{T_n}{n} \to \frac{1}{\lambda} \qquad \text{almost surely, as } n\to\infty.$$

Example 67 (Return times of Markov chains) For a positive-recurrent discrete-time Markov chain we denoted by

$$N_i = N_i^{(1)} = \inf\{n > 0 : M_n = i\}, \quad N_i^{(m+1)} = \inf\{n > N_i^{(m)} : M_n = i\}, m \in \mathbb{N},$$

the successive return times to 0. By the strong Markov property, the random variables $N_i^{(m+1)} - N_i^{(m)}$, $m \geq 1$ are independent and identically distributed. If we define $N_i^{(0)} = 0$ and start from i, then this holds for $m \geq 0$. The Strong Law of Large Number yields

$$\frac{N_i^{(m)}}{m} \to \mathbb{E}_i(N_i) \qquad \text{almost surely, as } m \to \infty.$$

Similarly, in continuous time, for

$$H_i = H_i^{(1)} = \inf\{t \geq T_1 : X_t = i\}, \quad H_i^{(m)} = T_{N_i^{(m)}}, m \in \mathbb{N},$$

we get

$$\frac{H_i^{(m)}}{m} \to \mathbb{E}_i(H_i) = m_i \qquad \text{almost surely, as } m \to \infty.$$

Example 68 (Empirical distributions) If $(Y_n)_{n \geq 1}$ is an infinite sample (independent and identically distributed random variables) from a discrete distribution ν on $\mathbb{S}$, then the random variables $B_n^{(i)} = 1_{\{Y_n = i\}}$, $n \geq 1$, are also independent and identically distributed for each fixed $i \in \mathbb{S}$, as functions of independent variables. The Strong Law of Large Numbers yields

$$\nu_i^{(n)} = \frac{\#\{k = 1, \ldots, n : Y_k = i\}}{n} = \frac{B_1^{(i)} + \ldots + B_n^{(i)}}{n} \to \mathbb{E}(B_1^{(i)}) = \mathbb{P}(Y_1 = i) = \nu_i$$

almost surely, as $n \to \infty$. The probability mass function $\nu^{(n)}$ is called empirical distribution. It lists relative frequencies in the sample and, for a specific realisation, can serve as an approximation of the true distribution. In applications of statistics, it is the sample distribution associated with a population distribution. The result that empirical distributions converge to the true distribution, is true uniformly in i and in higher generality, it is usually referred to as the Glivenko-Cantelli theorem.

Remark 69 (Discrete ergodic theorem) If $(M_n)_{n \geq 0}$ is a positive-recurrent discrete-time Markov chain, the Ergodic Theorem is a statement very similar to the example of empirical distributions

$$\frac{\#\{k = 0, \ldots, n-1 : M_k = i\}}{n} \to \mathbb{P}_\eta(M_0 = i) = \eta_i \qquad \text{almost surely, as } n \to \infty,$$

for a stationary distribution η, but of course, the M_n, $n \geq 0$, are not independent (in general). Therefore, we need to work a bit harder to deduce the Ergodic Theorem from the Strong Law of Large Numbers.

Lecture 10

Renewal processes and equations

Reading: Grimmett-Stirzaker 10.1-10.2; Ross 7.1-7.3

10.1 Motivation and definition

So far, the topic has been continuous-time Markov chains, and we've introduced them as discrete-time Markov chains with exponential holding times. In this setting we have a theory very much similar to the discrete-time theory, with independence of future and past given the present (Markov property), transition probabilities, invariant distributions, class structure, convergence to equilibrium, ergodic theorem, time reversal, detailed balance etc. A few odd features can occur, mainly due to explosion.

These parallels are due to the exponential holding times and their lack of memory property which is the key to the Markov property in continuous time. In practice, this assumption is often not reasonable.

Example 70 Suppose that you count the changing of batteries for an electrical device. Given that the battery has been in use for time t, is its residual lifetime distributed as its total lifetime? We would assume this, if we were modelling with a Poisson process.

We may wish to replace the exponential distribution by other distributions, e.g. one that cannot take arbitrarily large values or, for other applications, one that can produce clustering effects (many short holding times separated by significantly longer ones). We started the discussion of continuous-time Markov chains with birth processes as generalised Poisson processes. Similarly, we start here generalising the Poisson process to have non-exponential but independent identically distributed inter-arrival times.

Definition 71 Let $(Z_n)_{n\geq 0}$ be a sequence of independent identically distributed positive random variables, $T_n = \sum_{k=0}^{n-1} Z_k$, $n \geq 1$, the partial sums. Then the process $X = (X_t)_{t\geq 0}$ defined by

$$X_t = \#\{n \geq 1 : T_n \leq t\}$$

is called a *renewal process*. The common distribution of Z_n, $n \geq 0$, is called *inter-arrival distribution*.

Example 72 If $(Y_t)_{t\geq 0}$ is a continuous-time Markov chain with $Y_0 = i$, then $Z_n = H_i^{(n+1)} - H_i^{(n)}$, the times between successive returns to i by Y, are independent and identically distributed (by the strong Markov property). The associated counting process

$$X_t = \#\{n \geq 1 : H^{(n)} \leq t\}$$

counting the visits to i is thus a renewal process.

10.2 The renewal function

Definition 73 The function $t \mapsto m(t) := \mathbb{E}(X_t)$ is called the *renewal function.*

It plays an important role in renewal theory. Remember that for $Z_n \sim Exp(\lambda)$ we had $X_t \sim Poi(\lambda t)$ and in particular $m(t) = \mathbb{E}(X_t) = \lambda t$.

To calculate the renewal function for general renewal processes, we should investigate the distribution of X_t. Note that, as for birth processes,

$$X_t = k \iff T_k \leq t < T_{k+1},$$

so that we can express

$$\mathbb{P}(X_t = k) = \mathbb{P}(T_k \leq t < T_{k+1}) = \mathbb{P}(T_k \leq t) - \mathbb{P}(T_{k+1} \leq t)$$

in terms of the distributions of $T_k = Z_0 + \ldots + Z_{k-1}$, $k \geq 1$.

Recall that for two independent continuous random variables S and T with densities f and g, the random variable $S + T$ has density

$$(f * g)(u) = \int_{-\infty}^{\infty} f(u-t)g(t)dt, \quad u \in \mathbb{R},$$

the *convolution (product)* of f and g, and if $S \geq 0$ and $T \geq 0$, then

$$(f * g)(u) = \int_0^u f(u-t)g(t)dt, \quad u \geq 0.$$

It is not hard to check that the convolution product is symmetric, associative and distributes over sums of functions. While the first two of these properties translate as $S + T = T + S$ and $(S + T) + U = S + (T + U)$ for associated random variables, the third property has no such meaning, since sums of densities are no longer probability densities. However, the definition of the convolution product makes sense for general nonnegative integrable functions, and we will meet other relevant examples soon. We can define convolution powers $f^{*(1)} = f$ and $f^{*(k+1)} = f * f^{*(k)}$, $k \geq 1$. Then

$$\mathbb{P}(T_k \leq t) = \int_0^t f_{T_k}(s)ds = \int_0^t f^{*(k)}(s)ds,$$

if Z_n, $n \geq 0$, are continuous with density f.

Proposition 74 *Let X be a renewal process with interarrival density f. Then $m(t) = \mathbb{E}(X_t)$ is differentiable in the weak sense that it is the integral function of*

$$m'(s) := \sum_{k=1}^{\infty} f^{*(k)}(s)$$

Lemma 75 *Let X be an $\mathbb{N}$-valued random variable. Then $\mathbb{E}(X) = \sum_{k\geq 1} \mathbb{P}(X \geq k)$.*

Proof: We use Tonelli's Theorem

$$\sum_{k\geq 1} \mathbb{P}(X \geq k) = \sum_{k\geq 1}\sum_{j\geq k} \mathbb{P}(X = j) = \sum_{j\geq 1}\sum_{k=1}^{j} \mathbb{P}(X = j) = \sum_{j\geq 0} j\mathbb{P}(X = j) = \mathbb{E}(X).$$

□

Proof of Proposition 74: Let us integrate $\sum_{k=1}^{\infty} f^{*(k)}(s)$ using Tonelli's Theorem

$$\int_0^t \sum_{k=1}^{\infty} f^{*(k)}(s)ds = \sum_{k=1}^{\infty} \int_0^t f^{*(k)}(s)ds = \sum_{k=1}^{\infty} \mathbb{P}(T_k \leq t) = \sum_{k=1}^{\infty} \mathbb{P}(X_t \geq k) = \mathbb{E}(X_t) = m(t).$$

□

10.3 The renewal equation

For continuous-time Markov chains, conditioning on the first transition time was a powerful tool. We can do this here and get what is called the *renewal equation*.

Proposition 76 *Let X be a renewal process with interarrival density f. Then $m(t) = \mathbb{E}(X_t)$ is the unique (locally bounded) solution of*

$$m(t) = F(t) + \int_0^t m(t-s)f(s)ds, \qquad \text{i.e. } m = F + f * m,$$

where $F(t) = \int_0^t f(s)ds = \mathbb{P}(Z_1 \leq t)$.

Proof: Conditioning on the first arrival will involve the process $\widetilde{X}_u = X_{T_1+u}$, $u \geq 0$. Note that $\widetilde{X}_0 = 1$ and that $\widetilde{X}_u - 1$ is a renewal process with interarrival times $\widetilde{Z}_n = Z_{n+1}$, $n \geq 0$, independent of T_1. Therefore

$$\mathbb{E}(X_t) = \int_0^{\infty} f(s)\mathbb{E}(X_t|T_1 = s)ds = \int_0^t f(s)\mathbb{E}(\widetilde{X}_{t-s})ds = F(s) + \int_0^t f(s)m(t-s)ds.$$

For uniqueness, suppose that also $\ell = F + f * \ell$, then $\alpha = \ell - m$ is locally bounded and satisfies $\alpha = f * \alpha = \alpha * f$. Iteration gives $\alpha = \alpha * f^{*(k)}$ for all $k \geq 1$ and, summing over k gives for the right hand side something finite:

$$\begin{aligned} \left| \left(\sum_{k\geq 1} \alpha * f^{*(k)} \right)(t) \right| &= \left| \left(\alpha * \sum_{k\geq 1} f^{*(k)} \right)(t) \right| = \left| (\alpha * m')(t) \right| \\ &= \left| \int_0^t \alpha(t-s) m'(s) ds \right| \leq \left(\sup_{u\in[0,t]} |\alpha(u)| \right) m(t) < \infty \end{aligned}$$

but the left-hand side is infinite unless $\alpha(t) = 0$. Therefore $\ell(t) = m(t)$, for all $t \geq 0$. □

Example 77 We can express m as follows: $m = F + F * \sum_{k\geq 1} f^{*(k)}$. Indeed, we check that $\ell = F + F * \sum_{k\geq 1} f^{*(k)}$ satisfies the renewal equation:

$$F + f * \ell = F + F * f + F * \sum_{j\geq 2} f^{*(j)} = F + F * \sum_{k\geq 1} f^{*(k)} = \ell,$$

just using properties of the convolution product. By Proposition 76, $\ell = m$.

Unlike Poisson processes, general renewal processes do not have a linear renewal function, but it will be asymptotically linear (Elementary Renewal Theorem, as we will see). In fact, renewal functions are in one-to-one correspondence with interarrival distributions – we do not prove this, but it should not be too surprising given that $m = F + f * m$ is almost symmetric in f and m. Unlike the Poisson process, increments of general renewal processes are not stationary (unless we change the distribution of Z_0 in a clever way, as we will see) nor independent. Some of the important results in renewal theory are asymptotic results.

These asymptotic results will, in particular, allow us to prove the Ergodic Theorem for Markov chains.

10.4 Strong Law and Central Limit Theorem of renewal theory

Theorem 78 (Strong Law of renewal theory) *Let X be a renewal process with mean interarrival time $\mu \in (0, \infty)$. Then*

$$\frac{X_t}{t} \to \frac{1}{\mu} \qquad \text{almost surely, as } t \to \infty.$$

Proof: Note that X is constant on $[T_n, T_{n+1})$ for all $n \geq 0$, and therefore constant on $[T_{X_t}, T_{X_t+1}) \ni t$. Therefore, as soon as $X_t > 0$,

$$\frac{T_{X_t}}{X_t} \leq \frac{t}{X_t} < \frac{T_{X_t+1}}{X_t} = \frac{T_{X_t+1}}{X_t+1} \frac{X_t+1}{X_t}.$$

Now $\mathbb{P}(X_t \to \infty) = 1$, since $X_\infty \le n \iff T_{n+1} = \infty$ which is absurd, since $T_{n+1} = Z_0 + \ldots + Z_n$ is a finite sum of finite random variables. Therefore, we conclude from the Strong Law of Large Numbers for T_n, that

$$\frac{T_{X_t}}{X_t} \to \mu \qquad \text{almost surely, as } t \to \infty.$$

Therefore, if $X_t \to \infty$ and $T_n/n \to \mu$, then

$$\mu \le \lim_{t\to\infty} \frac{t}{X_t} \le \mu \qquad \text{as } t \to \infty,$$

but this means $\mathbb{P}(X_t/t \to 1/\mu) \ge \mathbb{P}(X_t \to \infty, T_n/n \to \mu) = 1$, as required. □

Try to do this proof for convergence in probability. The nasty ε expressions are not very useful in this context, and the proof is very much harder. But we can now deduce a corresponding Weak Law of Renewal Theory, because almost sure convergence implies convergence in probability.

We also have a Central Limit Theorem:

Theorem 79 (Central Limit Theorem of Renewal Theory) *Let $X = (X_t)_{t\ge0}$ be a renewal process whose interarrival times $(Y_n)_{n\ge0}$ satisfy $0 < \sigma^2 = Var(Y_1) < \infty$ and $\mu = \mathbb{E}(Y_1)$. Then*

$$\frac{X_t - t/\mu}{\sqrt{t\sigma^2/\mu^3}} \to \mathcal{N}(0,1) \qquad \textit{in distribution, as } t \to \infty.$$

The proof is not difficult and left as an exercise on Assignment 5.

10.5 The elementary renewal theorem

Theorem 80 *Let X be a renewal process with mean interarrival times μ and $m(t) = \mathbb{E}(X_t)$. Then*

$$\frac{m(t)}{t} = \frac{\mathbb{E}(X_t)}{t} \to \frac{1}{\mu} \qquad \textit{as } t \to \infty$$

Note that this does not follow easily from the strong law of renewal theory since almost sure convergence does not imply convergence of means (cf. Proposition 60, see also the counter example on Assignment 5). In fact, the proof is longer and not examinable: we start with a lemma.

Lemma 81 *For a renewal process X with arrival times $(T_n)_{n\ge1}$, we have*

$$\mathbb{E}(T_{X_t+1}) = \mu(m(t)+1), \qquad \textit{where } m(t) = \mathbb{E}(X_t),\ \mu = \mathbb{E}(T_1).$$

This ought to be true, because T_{X_t+1} is the sum of X_t+1 interarrival times, each with mean μ. Taking expectations, we should get $m(t)+1$ times μ. However, if we condition on X_t we have to know the distribution of the residual interarrival time after t, but without lack of memory, we are stuck.

Proof: Let us do a one-step analysis on the quantity of interest $g(t)=\mathbb{E}(T_{X_t+1})$:

$$g(t)=\int_0^\infty \mathbb{E}(T_{X_t+1}|T_1=s)f(s)ds=\int_0^t \left(s+\mathbb{E}(T_{X_{t-s}+1})\right)f(s)ds+\int_t^\infty sf(s)ds=\mu+(g*f)(t).$$

This is almost the renewal equation. In fact, $g_1(t)=g(t)/\mu-1$ satisfies the renewal equation

$$g_1(t)=\frac{1}{\mu}\int_0^t g(t-s)f(s)ds=\int_0^t(g_1(t-s)+1)f(s)ds=F(t)+(g_1*f)(t),$$

and, by Proposition 76, $g_1(t)=m(t)$, i.e. $g(t)=\mu(1+m(t))$ as required. □

Proof of Theorem 80: Clearly $t<\mathbb{E}(T_{X_t+1})=\mu(m(t)+1)$ gives the lower bound

$$\liminf_{t\to\infty}\frac{m(t)}{t}\geq\frac{1}{\mu}.$$

For the upper bound we use a truncation argument and introduce

$$\widetilde{Z}_j=Z_j\wedge a=\begin{cases} Z_j & \text{if } Z_j<a\\ a & \text{if } Z_j\geq a\end{cases}$$

with associated renewal process $\widetilde{X}$. $\widetilde{Z}_j\leq Z_j$ for all $j\geq 0$ implies $\widetilde{X}_t\geq X_t$ for all $t\geq 0$, hence $\widetilde{m}(t)\geq m(t)$. Putting things together, we get from the lemma again

$$t\geq\mathbb{E}(\widetilde{T}_{\widetilde{X}_t})=\mathbb{E}(\widetilde{T}_{\widetilde{X}_t+1})-\mathbb{E}(\widetilde{Z}_{\widetilde{X}_t+1})=\widetilde{\mu}(\widetilde{m}(t)+1)-\mathbb{E}(\widetilde{Z}_{\widetilde{X}_t+1})\geq\widetilde{\mu}(m(t)+1)-a.$$

Therefore

$$\frac{m(t)}{t}\leq\frac{1}{\widetilde{\mu}}+\frac{a-\widetilde{\mu}}{\widetilde{\mu}t}$$

so that

$$\limsup_{t\to\infty}\frac{m(t)}{t}\leq\frac{1}{\widetilde{\mu}}$$

Now $\widetilde{\mu}=\mathbb{E}(\widetilde{Z}_1)=\mathbb{E}(Z_1\wedge a)\to\mathbb{E}(Z_1)=\mu$ a $a\to\infty$ (by monotone convergence). Therefore

$$\limsup_{t\to\infty}\frac{m(t)}{t}\leq\frac{1}{\mu}.$$

□

Note that truncation was necessary to get $\mathbb{E}(\widetilde{Z}_{\widetilde{X}_t+1})\leq a$. It would have been enough if we had $\mathbb{E}(Z_{X_t+1})=\mathbb{E}(Z_1)=\mu$, but this is *not* true. Look at the Poisson process as an example. We know that the residual lifetime has already mean $\mu=1/\lambda$, but there is also the part of Z_{X_t+1} before time t. We will explore this in Lecture 11 when we discuss residual lifetimes in renewal theory.