Foundations of Stochastic Modeling

Lecture 1 – *19th January 2010*

Probability Preliminaries

- *Set operations*
  - When we talk of a probability space, we quote a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where
    - Ω is the space of outcomes
    - $\mathcal{F}$ is a σ-field – the space of "admissible sets/events". The sets we will be talking about henceforth reside in $\mathcal{F}$.
    - $\mathbb{P}$ is a probability distribution/measure based on which statements are made about the probability of various events. $\mathbb{P}:\mathcal{F}\to[0,1]$. In particular, for $A\in\mathcal{F}$, $\mathbb{P}(A)=\mathbb{P}(\omega\in\Omega:\omega\in A)$.
  - Note the following basic property of unions and intersections of sets

    $$\left[\bigcup_n A_n\right]^c = \bigcap_n A_n^c$$

    Or equivalently:

    $$\bigcup_n A_n^c = \left[\bigcap_n A_n\right]^c$$

  - When we say an event A holds "almost surely", we mean that
    - $\mathbb{P}(A)=1$
    - $A\in\mathcal{F}$ (in other words, A is *measurable*). We will assume this always hold throughout this course and omit this condition.
  - ***Proposition***: Let $\{A_n\}$ be a sequence of (measurable) events such that $\mathbb{P}(A_n)=1$ for all n. Then $\mathbb{P}\left(\bigcap_n A_n\right)=1$.

    Proof: First consider that

$$\left[\bigcap_n A_n\right]^c = \bigcup_n A_n^c$$

Then consider that, using the union bound

$$\begin{aligned}\mathbb{P}\left(\bigcup_n A_n^c\right) &\le \sum_n \mathbb{P}\left(A_n^c\right) \\ &= \sum_n 1 - \mathbb{P}\left(A_n\right) \\ &= \sum_n 0 \\ &= 0\end{aligned}$$

The union bound, used in the first line, simply states that if we add the probability of events without subtracting their intersections, the result will be greater than when considering the union (which leaves out intersections). ■

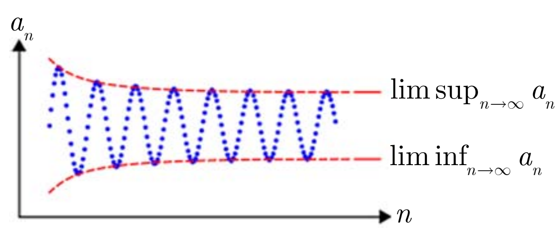
- ***Definition (limsup and liminf)***: Let $\{a_n\}$ be a real-value sequence. Then we write

$$\limsup_n a_n := \inf_m \sup_{n \ge m} a_n$$

This is the least upper bound on that sequence. Similarly

$$\liminf_n a_n := \sup_m \inf_{n \ge m} a_n$$

Graphically (source: Wikipedia):



Note that these limits need not be finite.

Remarks:

- If $\limsup_n a_n < a$, then $a_n < a$ *eventually* (ie: for sufficiently large n). If $\limsup_n a_n > a$, then $a_n > a$ infinitely often – in other words, there are infinitely many elements in the sequence above a.
- Similarly, $\liminf_n a_n > a \Rightarrow a_n > a$ ev. and $\liminf a_n < a \Rightarrow a_n < a$ i.o.

We now attempt to extend this definition to sets.

- ***Definition (limsup and liminf for sets)***: Let $\{A_n\}$ be a sequence of measurable sets. Then

$$\limsup_n A_n := \bigcap_m \bigcup_{n\geq m} A_n = \lim_{m\to\infty} \bigcup_{n\geq m} A_n$$
$$\liminf_n A_n := \bigcup_m \bigcap_{n\geq m} A_n = \lim_{m\to\infty} \bigcap_{n\geq m} A_n$$

We can interpret these as follows. The lim sup of A_n is the set of events that are contained in infinitely many of the A_n (but not necessarily in all the A_n past a point – the event could "bounce in and out" of the A_n)

$$\left\{\limsup_n A_n\right\} = \left\{\omega \in \Omega : \omega \in A_n \text{ i.o.}\right\}$$

Similarly, the lim inf of A_n is the set of events that eventually appear in an A_n and then in *all* A_n past that point.

$$\left\{\liminf_n A_n\right\} = \left\{\omega \in \Omega : \omega \in A_n \text{ ev.}\right\}$$

Clearly, if an event is in the lim inf, it also appears infinitely often (because it appears in all the A_n past a point, and so $\liminf_n A_n \subseteq \limsup_n A_n$.

Remark: Take $\left\{A_n, \text{ev.}\right\} = \left\{\liminf_n A_n\right\}$. Then

$$\begin{aligned}\left\{A_n, \text{ ev.}\right\}^c &= \left[\bigcup_m \bigcap_{n\geq m} A_n\right]^c \\ &= \bigcap_m \left\{\left(\bigcap_{n\geq m} A_n\right)^c\right\} \\ &= \bigcap_m \bigcup_{n\geq m} A_n^c \\ &= \liminf_n A_n^c \\ &= \left\{A_n^c, \text{ i.o.}\right\}\end{aligned}$$

- ***Proposition***: Let $\left\{A_n\right\}$ be a sequence of measurable sets, then
  - $\mathbb{P}\left(\liminf_n A_n\right) \leq \liminf_n \mathbb{P}\left(A_n\right)$
  - $\mathbb{P}\left(\limsup_n A_n\right) \geq \limsup_n \mathbb{P}\left(A_n\right)$

(The first statement can be thought of as a generalization of Fatou's Lemma for probabilities).

Proof of (i): Let us define $B_m = \bigcap_{n\geq m} A_n$. Then we know that $\cdots \subseteq B_m \subseteq B_{m+1} \subseteq \cdots$. In other words, the sets B_m increase monotonically to

$$B = \bigcup_m B_m = \bigcup_m \bigcap_{n\geq m} A_n = \liminf_n A_n$$

Since the events are increasing, a simple form of monotone convergence gives us that $\mathbb{P}\left(B_n\right) \nearrow \mathbb{P}\left(B\right)$. But we also have that

$$\mathbb{P}\left(B_m\right) \leq \mathbb{P}\left(A_n\right) \qquad n \geq m$$

because B_m is the intersection of all events from A_m onwards, so its probability is less than or equal to the probability of any single event. Thus

$$\begin{aligned} \mathbb{P}\left(B_m\right) &\leq \inf_{n\geq m} \mathbb{P}\left(A_n\right) \\ \sup_m \mathbb{P}\left(B_m\right) &\leq \sup_m \inf_{n\geq m} \mathbb{P}\left(A_n\right) \\ \mathbb{P}\left(B\right) &\leq \liminf_n \mathbb{P}\left(A_n\right) \end{aligned}$$

And given how we have defined B:

$$\mathbb{P}\left(\liminf_n A_n\right) \leq \liminf_n \mathbb{P}\left(A_n\right)$$

As required. ■

- ***Borel-Cantelli Lemmas & Independence***
  - ***Proposition (First Borel-Cantelli Lemma)*** Let $\left\{A_n\right\}$ be a sequence of measurable events such that $\sum_n \mathbb{P}(A_n) < \infty$. Then $\mathbb{P}\left(A_n \text{ i.o.}\right) = 0$. In other words, $\mathbb{P}\left(\limsup_n A_n\right) = 0$.[1] We offer two proofs – the first is somewhat mechanical, the second is more intuitive.

    Proof (version 1): Consider that

    $$\limsup_n A_n = \bigcap_m \bigcup_{n\geq m} A_n$$

    As such

    $$\begin{aligned} \mathbb{P}\left(\limsup_n A_n\right) &\leq \mathbb{P}\left(\bigcup_{n\geq m} A_n\right) \\ &\leq \sum_{n\geq m} \mathbb{P}\left(A_n\right) \\ &\to 0 \end{aligned}$$

    As required. ■

    The second proof will require a lemma:

    Lemma (Fubini's): If $\left\{X_n\right\}$ is a sequence of real-valued random variables with $X_n \geq 0$, then $\mathbb{E}\left(\sum_n X_n\right) = \sum_n \mathbb{E}\left(X_n\right)$ (which could be infinite). This is

[1] It might not be clear that the statements $\mathbb{P}\left(A_n \text{ i.o.}\right) = 0$ and $\mathbb{P}\left(\limsup_n A_n\right) = 0$ are equivalent. To see this more clearly, note that the first statement is simply a shorthand for $\mathbb{P}\left(\omega \in \Omega : \omega \in A_n \text{ i.o.}\right) = 0$. This is clearly identical, by definition, to $\mathbb{P}\left(\limsup_n A_n\right) = 0$.

effectively a condition under which we can exchange a summation and an integration.

Proof (version 2): Let

$$X_n(\omega) = \mathbb{I}_{A_n} = \begin{cases} 1 & \text{if } \omega \in A_n \\ 0 & \text{otherwise} \end{cases}$$

Note that $\mathbb{E}(X_n) = \mathbb{P}(A_n)$. We also have that

$$\begin{aligned} \sum_n \mathbb{P}(A_n) &= \sum_n \mathbb{E}(\mathbb{I}_{A_n}) \\ &= \mathbb{E}\left(\sum_n \mathbb{I}_{A_n}\right) \\ &< \infty \end{aligned}$$

This means that the random variable $\sum_n \mathbb{I}_{A_n}$ must be less than or equal to infinity for every outcome in Ω; in other words

$$\sum_n \mathbb{I}_{A_n} < \infty \qquad \text{a.s.}$$

This must mean that $\mathbb{I}_{A_n} = 1$ only finitely many times. Thus, A_n occurs only finitely many times. ■

- ***Definition (independence)***: A sequence of random variables $\{X_1, \cdots, X_n\}$ are independent if and only if

  $$\mathbb{P}(X_1 \le x_1, \cdots, X_n \le x_n) = \prod_{i=1}^n \mathbb{P}(X_i \le x_i)$$

  For all $x_1, \cdots, x_n \in \mathbb{R}$.

  Defininition (independence): A sequence of sets $\{A_1, \cdots, A_n\}$ are said to be *independent* if and only if the sequence or random variables $\{\mathbb{I}_{A_i}, \cdots, \mathbb{I}_{A_n}\}$ are independent.

- ***Proposition (second Borrel-Cantelli Lemma)***: Let $\{A_n\}$ be a sequence of independent measurable events such that $\sum_n \mathbb{P}(A_n) = \infty$, then $\mathbb{P}(A_n \text{ i.o.}) = 1$.

  Proof: We showed in a remark above that

  $$\begin{aligned} \{A_n, \text{i.o}\}^c &= \{A_n^c, \text{ev.}\} \\ &= \bigcup_m \bigcap_{n \ge m} A_n^c \end{aligned}$$

  Let us fix $m \ge 1$. Then using the inequality $1 - x \le e^{-x}$

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

But we have that

$$\begin{aligned}\mathbb{P}\left(\bigcup_m \bigcap_{n\geq m} A_n^c\right) &\leq \sum_m \mathbb{P}\left(\bigcap_{n\geq m} A_n^c\right) \\ &= 0\end{aligned}$$

As required. ■

- ***Notions of convergence***
  - ***Definition (convergence almost surely)*** $X_n \to X$ almost surely if

    $$\mathbb{P}\left\{\omega \in \Omega : X_n(\omega) \to X(\omega)\right\} = 1$$

    More generally, $\left\{X_n\right\}$ converges almost surely if

    $$\mathbb{P}\left(\limsup_n X_n = \liminf_n X_n\right) = 1$$

    This is the notion of convergence that most closely maps to convergence concepts in real analysis.

    Let us understand these two statements intuitively

    - The first statement tells us that for *every single outcome* in the sample space Ω, the sequence of random variables X_n tends to X.
    - The second statement is a shorthand for the statement that

      $$\mathbb{P}\left(\left\{\omega : \limsup_n X_n(\omega) = \liminf_n X_n(\omega)\right\}\right) = 1$$

      This is, again, simply the statement in real analysis that

      $$\limsup_n x_n = \liminf_n x_n \text{ if and only if } \lim_n x_n \text{ exists}$$

      Thus, we require the limit to exist for *every outcome* ω.
  - ***Example***: Consider the sequence $X_n = \frac{1}{n} U[0,1]$. We claim that $X_n \to 0$ almost surely.

    Proof: In this case, $\Omega = [0,1]$. For any ω we might drawn, we will find

    $$0 \leq X_n(\omega) \leq \tfrac{1}{n} \to 0$$

    As required. ■ □

- ***Definition (convergence in probability)***: $X_n \to_p X$ as $n \to \infty$ if, for all $\varepsilon > 0$

$$\mathbb{P}\left(\left|X_n - X\right| > \varepsilon\right) \to 0 \qquad \text{as } n \to \infty$$

Let us understand this concept as compared to convergence almost surely.

  - In this case, we consider all the outcomes for which X_n is far from X, and we require the *probability* of these outcomes to get smaller as n progresses. This, however, allows the possibility that for some outcomes ω, $X_n(\omega) \not\to X(\omega)$ infinitely often, provided that the probability of these ω is small and gets smaller.
  - In the case of almost sure convergence, we only allow the deviation to be large a finite number of times. For *every* outcome, we require $X_n(\omega) \to X(\omega)$. Perhaps this appears clearer if we write the condition for convergence almost surely as

$$\mathbb{P}\left(\left|X_n - X\right| > \varepsilon \ \text{ i.o.}\right) = 0$$

The difference between the two statements above really hinges on the novelty of the BC Lemmas. If A_n is the event that "X_n is far from X", convergence in probability requires $\lim_{n\to\infty} \mathbb{P}(A_n) = 0$, whereas convergence almost surely requires that $\mathbb{P}(A_n \text{ i.o.}) = 0$ – which would, for example, be satisfied if $\sum_n \mathbb{P}(A_n) < \infty$.

- Another interesting way to view the difference between convergence almost surely and in probability is in terms of experiments. Imagine carrying out an "experiment" by generating random variables $X_1, X_2, \cdots$. Each such experiment corresponds to a different outcome $\omega \in \Omega$.
  - If $X_n \to X$ almost surely, then it is *certain* that in *every* experiment we might carry out, the values $X_1(\omega), X_2(\omega), \cdots$ will tend to $X(\omega)$.
  - If $X_n \to_p X$, then it is possible that in some experiments (for a select subset of Ω), $X_1(\omega), X_2(\omega), \cdots$ will not tend to $X(\omega)$. The probability of

this happening decreases with the number of X generated, but it is still possible that in certain experiments, it will happen.

- ***Example***: Let $X_n = \mathbb{I}_{\{U_n \le \frac{1}{n}\}}$, where the U_n are IID $U[0, 1]$ random variables. Let us fix $\varepsilon \in (0,1)$, and consider the definition

$$\begin{aligned}\mathbb{P}\left(\left|X_n - 0\right| > \varepsilon\right) &= \mathbb{P}\left(X_n > \varepsilon\right) \\ &= \mathbb{P}\left(U_n \le \tfrac{1}{n}\right) \\ &= \frac{1}{n} \\ &\to 0\end{aligned}$$

This proves the X_n do indeed converge to 0 in probability. Consider, however, that

$$\sum_n \mathbb{P}\left(X_n > \varepsilon\right) = \sum_n \tfrac{1}{n} = \infty$$

But since our X_n are independent, we know by the second BC Lemma that $\left\{X_n > \varepsilon, \text{ i.o.}\right\}$ almost surely. So X_n cannot be converging to 0 almost surely. Note, on the other hand, that if we replaced our $1/n$ by $1/n^2$, we would find (using the first Borel-Cantelli Lemma) that X_n does indeed converge to 0 almost surely. □

- ***Claim***: $X_n \to X$ a.s. $\Rightarrow X_n \to_p X$ as $n \to \infty$. As our example shows, however, the converse is false.

Proof: Recall that convergence almost surely can be written as

$$\mathbb{P}\left(\left|X_n - X\right| > \varepsilon \quad \text{i.o.}\right) = 0$$

or in other words

$$\mathbb{P}\left(\limsup_n \left\{\left|X_n - X\right| > \varepsilon\right\}\right) = 0$$

We have, however, that $\liminf_n A_n \subseteq \limsup_n A_n$. As such

$$\begin{aligned}\mathbb{P}\left(\liminf_n \left\{\left|X_n - X\right| > \varepsilon\right\}\right) &= 0 \\ \mathbb{P}\left(\left|X_n - X\right| > \varepsilon \text{ e.v.}\right) &= 0\end{aligned}$$

This naturally implies that *eventually* the probability of large deviations falls to 0 – this is the definition of convergence in probability. ■

- ***Remark***: It is interesting to note that if $X_n \to_p X$, it is possible to find a *subsequence* $X_{n_k} \to X$ a.s., $k = 1, 2, \cdots$. This solidifies our intuition that convergence in probability differs from convergence almost surely only as a result of a few "freak outcomes".

  Proof: Assume we have X_{n_k}. Pick n_{k+1} to be the smallest n such that

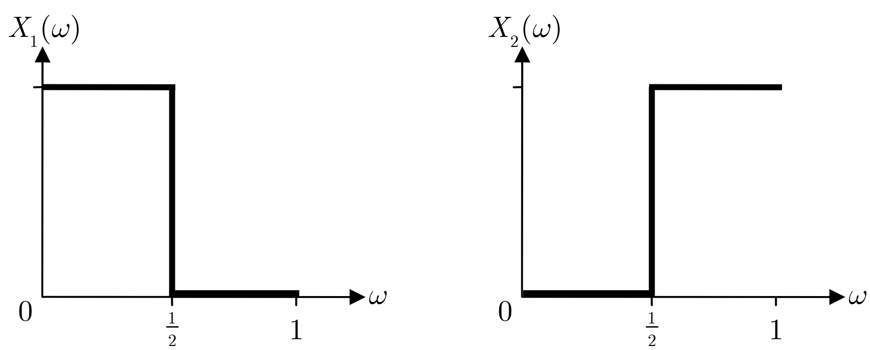
  $$\mathbb{P}\left(\left|X_{n_{k+1}} - X\right| \geq \tfrac{1}{k}\right) \leq \tfrac{1}{2^k}$$

  (This is always possible by the definition of convergence in probability). The convergence of X_{n_k} almost surely follows by the BC Lemma. ■

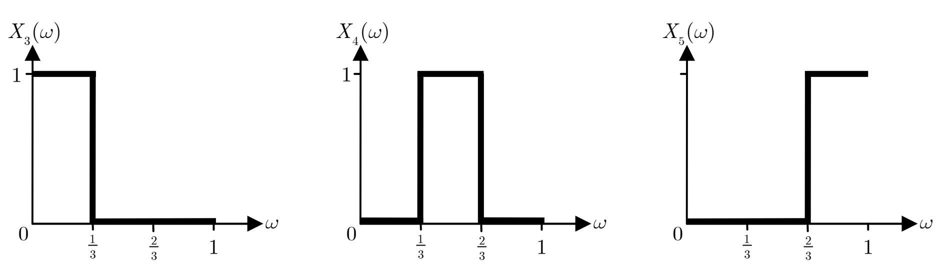
- ***Example***: Take $\Omega = [0,1]$ and $\mathbb{P} =$ uniform distribution. We then define random variables X_1 and X_2 as follows:

  $$X_1(\omega) = \begin{cases} 1 & \omega \in [0, \tfrac{1}{2}) \\ 0 & \text{otherwise} \end{cases} \qquad X_2(\omega) = \begin{cases} 0 & \omega \in [0, \tfrac{1}{2}) \\ 1 & \text{otherwise} \end{cases}$$

  Graphically:



  Similarly, we define X_3, X_4 and X_5 to be random variables divided into three parts:



  We continue this pattern to form a pyramid of random variables.

Now, consider $\varepsilon \in (0,1)$ – since X_n is equal to 0 or 1, saying $X_n > \varepsilon$ is the same as saying $X_n = 1$. Now, consider that $\mathbb{P}\left(X_1 = 1\right) = \mathbb{P}\left(X_2 = 1\right) = \frac{1}{2}$. Similarly, for random variables divided into three parts, the probability is $\frac{1}{3}$, etc… As such

$$\mathbb{P}\left(X_n > \varepsilon\right) \to 0 \qquad \text{as } n \to \infty$$

And so $X_n \to_p 0$.

However, if we fix any $\omega \in [0,1]$, $X_n(\omega) = 1$ infinitely often (because every set of random variables will involve at least one that is equal to 1 at that ω). Thus, X_n does not tend to 0 almost surely.

- ***Definition (convergence in expectation)*** – also called L_1 convergence. We say that $X_n \to_{L_1} X$ as $n \to \infty$ if

$$\mathbb{E}\left|X_n - X\right| \to 0 \qquad \text{as } n \to \infty$$

Similarly, $X_n \to_{L_p} X$ if

$$\mathbb{E}\left|X_n - X\right|^p \to 0 \qquad \text{as } n \to \infty$$

- ***Claim***: $X_n \to_{L_1} X$ implies that $\mathbb{E}\left|X_n\right| \to \mathbb{E}\left|X\right|$ and $\mathbb{E}(X_n) \to \mathbb{E}(X)$.

  Proof: For the first part, consider that

$$\begin{gathered}\left|X_n\right| = \left|X_n - X + X\right| \leq \left|X_n - X\right| + \left|X\right| \\ \Rightarrow \left|X_n\right| - \left|X\right| \leq \left|X_n - X\right|\end{gathered}$$

Taking expectations yields the result. For the second part, note that

$$\left|\mathbb{E}(X_n - X)\right| = \left|\mathbb{E}(X_n) - \mathbb{E}(X)\right| \leq \mathbb{E}\left(\left|X_n - X\right|\right)$$

Where the last step follows by Jensen's Inequality. Taking limits yields the answer required. ■

- ***Theorem (Markov's Inequality)***: Let $X \geq 0$ have finite expectation. Then for all $a > 0$

$$\mathbb{P}\left(X > a\right) \leq \frac{\mathbb{E}(X)}{a}$$

Proof:

$$\begin{aligned}\mathbb{E}(X) &= \mathbb{E}\left(X, X > a\right) + \mathbb{E}\left(X, X \le a\right) \\ &\ge \mathbb{E}\left(X, X > a\right) \\ &\ge a\mathbb{P}\left(X > a\right)\end{aligned}$$

As required. ■

- ***Claim***: If $X_n \to_{L_1} X$ then $X_n \to_p X$.

  Proof: Fix an $\varepsilon > 0$ and use Markov's Inequality

  $$\mathbb{P}\left(\left|X_n - X\right| > \varepsilon\right) \le \frac{\mathbb{E}\left(\left|X_n - X\right|\right)}{\varepsilon} \to 0$$

  As required. ■

- ***Example***: Consider $X_n = n\mathbb{I}_{\left\{U \le \frac{1}{n}\right\}}$, where U is $U[0,1]$. Clearly, $X_n \to 0$ a.s. (Note that this does *not* hold true for $\omega = 0$. However, this set has measure 0, and so the convergence is still *almost* sure). However, it is clear that $\mathbb{E}\left(X_n\right) = n\mathbb{P}\left(U \le \frac{1}{n}\right) = 1$. Clearly, therefore, L_1 convergence and almost-sure convergence are not compatible. (This illustrates an interesting example where taking limits inside and outside expectations gives a different result. Indeed, $1 = \lim_n \mathbb{E}(X_n) \ne \mathbb{E}\left(\lim_n X_n\right) = 0$.)

- ***Example***: Let X_n be a sequence of IID random variables with exponential distribution with parameter 1. We will prove that

  $$\limsup_n \frac{X_n}{\log n} = 1 \qquad \text{a.s.}$$

  Intuitively, this implies that if we simulate a certain number of exponentially distributed random variables and plot them against their index, and then plot the curve log n, an infinite number of these points will lie on the curve! This is quite shocking, since one might expect very few very large values!

  Proof: First $\varepsilon > 0$. Consider

  $$\begin{aligned}\mathbb{P}\left(X_n > (1+\varepsilon)\log n\right) &= e^{-(1+\varepsilon)\log n} \\ &= \frac{1}{n^{1+\varepsilon}}\end{aligned}$$

  This means that

  $$\sum_n \mathbb{P}\left(X_n > (1+\varepsilon)\log n\right) < \infty$$

So by BC-1,

$$\mathbb{P}\left(X_n > (1+\varepsilon)\log n,\ \text{i.o.}\right) = 0$$

As such,

$$X_n \le (1+\varepsilon)\log n \ \ \text{ev. almost surely}$$

and therefore

$$\boxed{\limsup_n \frac{X_n}{\log n} \le 1+\varepsilon}$$

By the same reasoning

$$\mathbb{P}\left(X_n > (1-\varepsilon)\log n\right) = \frac{1}{n^{1-\varepsilon}}$$
$$\Rightarrow \sum_n \mathbb{P}\left(X_n > (1-\varepsilon)\log n\right) = \infty$$

So by BC-2 and the independence of the X_n,

$$\mathbb{P}\left(X_n > (1-\varepsilon)\log n,\ \text{i.o.}\right) = 1$$
$$\Rightarrow \boxed{\limsup_n \frac{X_n}{\log n} > 1-\varepsilon \quad \text{a.s.}}$$

The two boxed results together imply our result. ■

Note: This only applies to the lim sup – to say that the actual sequence tends to 1 is nonsensical, because most of the X_n will indeed be very small!

LECTURE 2 – *26th January 2010*

- ## *Interchange arguments*
  - ***Example***: In our last lecture, we considered the sequence of random variables $X_n = n\mathbb{I}_{\{U \le \frac{1}{n}\}}$. We argued that $X_n \to 0$ almost surely (as $n \to \infty$). We showed, however, that $\mathbb{E}(X_n) = 1$ for all n. We might wonder whether it is generally true that

$$\mathbb{E}\left(\lim_n X_n\right) = \mathbb{E}\left(X\right) \stackrel{?}{=} \lim_n \mathbb{E}(X_n)$$

Clearly, this is not the case here. It looks like certain conditions need to be satisfied. □

- ***Theorem (BDD – Bounded Convergence)***: Let $\{X_n\}$ be a sequence of random variables such that $|X_n| \le K \le \infty$ (where K is a deterministic constant) and $X_n \to X$ almost surely, then $\mathbb{E}(X_n) \to \mathbb{E}(X)$. (Clearly, in the example above, the n that sits outside the indicator prevents us from bounding X_n).

  Proof: For any given $\varepsilon > 0$, let

  $$A_n = \left\{\omega \in \Omega : \left|X_n(\omega) - X(\omega)\right| < \varepsilon\right\}$$

  Let $B_n = \Omega - A_n$. Because $X_n \to X$ a.s., $\mathbb{P}(B_n) \to 0$. Thus

  $$\begin{aligned}\left|\mathbb{E}(X_n) - \mathbb{E}(X)\right| &\overbrace{\le}^{\text{Jensen's}} \mathbb{E}\left(\left|X_n - X\right|\right) \\ &= \int_{A_n} \left|X_n - X\right| \mathbb{P}(\mathrm{d}\omega) + \int_{B_n} \left|X_n - X\right| \mathbb{P}(\mathrm{d}\omega) \\ &\le \varepsilon \mathbb{P}(A_n) + 2K\mathbb{P}(B_n) \\ &\to 0\end{aligned}$$

  In the last line, we used the fact that for all outcomes in A_n, the absolute difference is less than ε (by definition) and that for all outcomes in B_n, the absolute difference is less than $2K$ (because $|X_n| \le K \Rightarrow |\lim_n X_n| = |X| \le K$. ■

  Remark: It is clear that this theorem still holds if X_n only converges *in probability* to X rather than almost surely, because all that is required is for the *probability* of rogue events to fall to 0. The fact that $|X| \le K$ can be deduced from the fact there is a subsequence X_{n_k} that converges to X almost surely (see above).

- ***Lemma (Fatou)***: If $\{X_n\}$ is a sequence of non-negative random variables, then

  $$\mathbb{E}\left(\liminf_n X_n\right) \le \liminf_n \mathbb{E}(X_n)$$

  (Note: there is no converse for lim sup).

  Proof: Let $Y_m = \inf_{n \ge m} X_n$ and $Y = \lim_n Y_n = \liminf_n X_n$. Note that:

  - $Y_m \to Y$ a.s.
  - $\inf_{n \ge m} \mathbb{E}(X_n) \ge \mathbb{E}(\inf_{n \ge m} X_n) = \mathbb{E}(Y_m)$. To see why, consider the X_n that has lowest expectation – call it $X_{\min}$. The LHS then has value $\int X_{\min}(\omega)\mathbb{P}(\mathrm{d}\omega)$. The RHS, however, has value $\int \inf_n X_n(\omega)\mathbb{P}(\mathrm{d}\omega)$. This is clearly smaller or equal to the LHS.

Now, pick some k

$$\begin{aligned}\liminf X_n &\geq \lim_{n\to\infty} \mathbb{E}\left(Y_m\right) \\ &\geq \lim_{n\to\infty} \mathbb{E}\left(\min[Y_m, k]\right)\end{aligned}$$

The absolute value of the variable of which we are taking an expectation is now bounded by k, because the X_n are positive, and so Y_m is positive. So we can apply bounded convergence:

$$\begin{aligned}\liminf X_n &\geq \mathbb{E}\left(\lim_{n\to\infty} \min[Y_m, k]\right) \\ &= \mathbb{E}\left(\min[Y, k]\right)\end{aligned}$$

As $k \to \infty$, the last line tends to $\mathbb{E}(Y) = \mathbb{E}(\liminf_n X_n)$. ■

- ***Theorem (DOM – Dominated Convergence)***: Let $\left\{X_n\right\}$ be a sequence of random variables such that $\left|X_n\right| \leq Y$ with $\mathbb{E}(Y) < \infty$ and $X_n \to X$ almost surely, then $\mathbb{E}(X_n) \to \mathbb{E}(X)$.

  Proof: Since $\left|X_n\right| \leq Y$, we have that $Y + X_n \geq 0$ and $Y - X_n \geq 0$. Applying Fatou's Lemma to both variables, we obtain

  $$\liminf_n \mathbb{E}\left(Y + X_n\right) \geq \mathbb{E}\left(Y + X\right) \qquad \liminf_n \mathbb{E}\left(Y - X_n\right) \geq \mathbb{E}\left(Y - X\right)$$

  Since Y has finite expectation, we can subtract $\mathbb{E}(Y)$ from both sides above, and obtain

  $$\begin{aligned}\liminf_n \mathbb{E}\left(X_n\right) &\geq \mathbb{E}\left(X\right) \qquad & \liminf_n \mathbb{E}\left(-X_n\right) &\geq \mathbb{E}\left(-X\right) \\ \liminf_n \mathbb{E}\left(X_n\right) &\geq \mathbb{E}\left(X\right) \qquad & \limsup_n \mathbb{E}\left(X_n\right) &\leq \mathbb{E}\left(X\right)\end{aligned}$$

  Together, the last two lines imply that $\lim_n \mathbb{E}(X_n) \to \mathbb{E}(X)$. ■

- ***Theorem (MON – Monotone Convergence)***: Let $\left\{X_n\right\}$ be a sequence of random variables such that $0 \leq X_1 \leq X_2 \leq \cdots$ almost surely, then
  - $X_n \nearrow X$ (possibly ∞) almost surely.
  - $\mathbb{E}(X_n) \nearrow \mathbb{E}(X)$ (possibly ∞).

  Proof: Since $X_n \geq 0$, we can apply Fatou's Lemma, to get

  $$\liminf_n \mathbb{E}(X_n) \geq \mathbb{E}(\liminf_n X) = \mathbb{E}(X)$$

  However, the fact that $X_n \leq X$ also gives $\mathbb{E}(X_n) \leq \mathbb{E}(X) \Rightarrow \liminf_n \mathbb{E}(X_n) \leq \mathbb{E}(X)$. These two statements together imply that $\lim_n \mathbb{E}(X_n) \to \mathbb{E}(X)$. ■

- ***Proposition (Frobini I)***: Let $\left\{X_n\right\}$ be a sequence of random variables with $X_n \geq 0$ for all n, then

$$\mathbb{E}\left(\sum_n X_n\right) = \sum_n \mathbb{E}(X_n)$$

Proof: Define

$$S_n = \sum_{m=1}^n X_n \nearrow S = \sum_{m=1}^\infty X_m \qquad \text{(possibly } \infty\text{)}$$

(The last step follows because the X_n are non-negative, which means the sequence $\left\{S_n\right\}$ is non-decreasing almost surely – there is no outcome in the probability space in which adding an extra term to the sum decreases it). By monotone convergence, we have that $\mathbb{E}(S_n) \nearrow \mathbb{E}(S)$. However,

$$\mathbb{E}\left(S_n\right) = \mathbb{E}\left(\sum_{m=1}^n X_n\right) \overbrace{=}^{\text{Finite } n} \sum_{m=1}^n \mathbb{E}(X_n) \nearrow \sum_{m=1}^\infty \mathbb{E}(X_n)$$

And

$$\mathbb{E}(S) = \mathbb{E}\left(\sum_{m=1}^\infty X_m\right)$$

Combining these two results yields

$$\sum_{m=1}^\infty \mathbb{E}(X_n) = \mathbb{E}\left(\sum_{m=1}^\infty X_m\right)$$

As required. ■

- ***Proposition (Frobini II)***: Let $\left\{X_n\right\}$ be a sequence of random variables with $\mathbb{E}\left(\sum_n \left|X_n\right|\right) < \infty$. Then

$$\mathbb{E}\left(\sum_n X_n\right) = \sum_n \mathbb{E}(X_n)$$

Proof: As above, set $S_n = \sum_{m=1}^n X_m$. By the Triangle inequality, we have

$$\left|S_n\right| = \left|\sum_{m=1}^n X_m\right| \leq \sum_{m=1}^n \left|X_m\right| \leq \overbrace{\sum_{m=1}^\infty \left|X_m\right|}^{Y} < \infty \qquad \text{a.s.}$$

(The last inequality follows because the statement of the Theorem includes the fact that the expectation of the infinite sum is smaller than infinity. If the sum was equal to infinity of a set of non-zero measure, the expectation would blow up. Thus, the sum must be is less than or equal to infinity almost surely.)

Now, note that

$$\begin{aligned}\left|S_n - \sum_{m=1}^{\infty} X_m\right| &= \left|\sum_{m=1}^{n} X_m - \sum_{m=1}^{\infty} X_m\right| \\ &= \left|\sum_{m=n+1}^{\infty} X_n\right| \\ &\leq \sum_{m=n+1}^{\infty} \left|X_m\right| \\ &\to 0 \qquad \text{a.s. as } n \to \infty\end{aligned}$$

(The last line follows from the fact the sum is finite almost surely. Thus, if we continually whittle away at the sum from below, it will eventually shrink to 0 almost surely).

The above implies that $S_n \to \sum_{m=1}^{\infty} X_m$ almost surely. We can now apply dominated convergence. First note that

$$\sum_{m=1}^{n} \mathbb{E}(X_m) \overbrace{=}^{\text{Finite } n} \mathbb{E}\left(\sum_{m=1}^{n} X_m\right) = \mathbb{E}(S_n)$$

Now take limits of both sides

$$\lim_n \sum_{m=1}^{n} \mathbb{E}(X_m) = \sum_{m=1}^{\infty} \mathbb{E}(X_m) = \lim_n \mathbb{E}(S_n)$$

By dominated convergence, we have

$$\lim_n \mathbb{E}(S_n) = \mathbb{E}(\lim_n S_n) = \mathbb{E}\left(\sum_{m=1}^{\infty} X_m\right)$$

Combining the last two equations, we do indeed find that

$$\sum_{m=1}^{\infty} \mathbb{E}(X_m) = \mathbb{E}\left(\sum_{m=1}^{\infty} X_m\right)$$

As required. ■

- We have, so far, looked at sufficient conditions for interchange. Is there a "common thread" that runs through these conditions?
- ***Definition (Uniform Integrability)***: A sequence of random variables $\{X_n\}$ is said to be *uniformly integrable* (u.i.) if for all $\varepsilon > 0$, there exists a $K(\varepsilon) < \infty$ such that

$$\sup_n \mathbb{E}\left(\left|X_n\right| : \left|X_n\right| > K(\varepsilon)\right) \equiv \sup_n \mathbb{E}\left(\left|X_n\right| \mathbb{I}_{\{|X_n| > K(\varepsilon)\}}\right) \leq \varepsilon$$

Note that the supremum implies that the K we find must work *uniformly* for the entire sequence – in other words, for a given ε, there must be a *single* $K(\varepsilon)$ which makes the tail expectation small for *every* n. Requiring this to be true for a *single* X_n is actually trivial, provided the expectation is finite, because a finite expectation implies that the tail eventually "dies out".

- ***Proposition***: Let $\{X_n\}$ be a sequence of random variables and suppose $X_n \to X$ almost surely
  - If $\{X_n\}$ are uniformly integrable, then $\mathbb{E}(X_n) \to \mathbb{E}(X)$. In fact, a stronger result applies – that $\mathbb{E}\left(\left|X_n - X\right|\right) \to 0$.
  - If $X_n > 0$ for all n and $\mathbb{E}(X_n) \to \mathbb{E}(X) < \infty$, then $\{X_n\}$ are uniformly integrable.

In the absence of the non-negativity requirement in the second statement, this would be an "if and only if" statement. We have almost identified the "bare minimum" condition for interchange.

Proof:

- ***Part 1***: First, let us verify that X is integrable

$$\begin{aligned}\mathbb{E}\left|X\right| &= \mathbb{E}\left(\liminf_n \left|X_n\right|\right) \\ &\overbrace{\le}^{\text{Fatou}} \liminf_n \mathbb{E}\left|X_n\right| \\ &\le \sup_n \mathbb{E}\left|X_n\right| \\ &= \sup_n \left(\mathbb{E}\left(\left|X_n\right|; \left|X_n\right| \le a\right) + \mathbb{E}\left(\left|X_n\right|; \left|X_n > a\right|\right)\right) \\ &\overbrace{\le}^{X_n \text{ are u.i}} \sup_n \left(a + \varepsilon\right) \\ &< \infty\end{aligned}$$

Now, let $Y_n = \left|X_n - X\right|$. We then have $0 \le Y_n \le \left|X_n\right| + \left|X\right|$. Hence, $\{Y_n\}$ is also uniformly integrable (since $\{X_n\}$ is uniformly integrable, and X is integrable). Write

$$\begin{aligned}\mathbb{E}(Y_n) &= \mathbb{E}\left(Y_n; Y_n > a\right) + \mathbb{E}\left(Y_n; Y_n \le a\right) \\ &\le \varepsilon + \mathbb{E}\left(Y_n \mathbb{I}_{\{Y_n \le a\}}\right)\end{aligned}$$

Now, note that $Y_n \mathbb{I}_{\{Y_n \le a\}} \le a$, so by bounded convergence – and since $Y_n \to 0$ a.s. because $X_n \to X$ a.s. – we have that $\mathbb{E}\left(Y_n \mathbb{I}_{\{Y_n \le a\}}\right) \to 0$. Thus, $\mathbb{E}(Y_n) \to 0$, as required.

- ***Part 2***: Given $a > 1$, define

$$f_a(x) = \begin{cases} x & x \in [0, a-1] \\ (\text{line connecting } a-1 \text{ to } 0) & x \in [a-1, a] \\ 0 & x > a \end{cases}$$

Clearly, f_a is continuous, and

$$x\mathbb{I}_{\{x \le a-1\}} \le f_a(x) \le x\mathbb{I}_{\{x \le a\}}$$

Now

$$\begin{aligned} \mathbb{E}\left(\left|X_n\right|\mathbb{I}_{\{|X_n|>a\}}\right) &= \mathbb{E}\left(\left|X_n\right|\right) - \mathbb{E}\left(\left|X_n\right|\mathbb{I}_{\{|X_n|\le a\}}\right) \\ &\le \mathbb{E}\left(\left|X_n\right|\right) - \mathbb{E}\left(f_a\left[\left|X_n\right|\right]\right) \end{aligned}$$

Now, f_a is clearly bounded, and $X_n \to X$. Thus, applying bounded convergence, $f_a\left(\mathbb{E}\left[\left|X_n\right|\right]\right)$. Furthermore, by the statement of the theorem, we have $\mathbb{E}\left(\left|X_n\right|\right) \to \mathbb{E}\left(\left|X\right|\right)$ (since the X_n are positive and $\mathbb{E}(X_n) \to \mathbb{E}(X)$). Thus, the RHS above can be made arbitrary close to the difference between these two expectations. In other words, there exists an n_0 such that for all $n > n_0$,

$$\begin{aligned} \mathbb{E}\left(\left|X_n\right|\mathbb{I}_{\{|X_n|>a\}}\right) &\le \mathbb{E}\left(\left|X_n\right|\right) - f_a\left(\mathbb{E}\left[\left|X\right|\right]\right) + \varepsilon \\ &\le \mathbb{E}\left(\left|X_n\right|\right) - \mathbb{E}\left(\left|X\right|\mathbb{I}_{\{|X|\le a-1\}}\right) + \varepsilon \end{aligned}$$

- ***Proposition (Sufficient conditions for u.i.)***
  - If $\sup_n \mathbb{E}\left|X_n\right|^p < \infty$ for some $p > 1$, then $\left\{X_n\right\}$ is uniformly integrable.
  - If $\left|X_n\right| \le Y_n$ and $\left\{Y_n\right\}$ is uniformly integrable, then $\left\{X_n\right\}$ is uniformly integrable.
  - If $\left\{X_n\right\}$ and $\left\{Y_n\right\}$ are uniformly integrable, then $\left\{X_n + Y_n\right\}$ is uniformly integrable.

Proof (part 1): We will prove the first part of the statement above. To do that, however, we will need to use *Holder's Inequality*.

Proposition (Holder's Inequality): Let X and Y be random variables such that $\mathbb{E}\left|X\right|^p < \infty$ and $\mathbb{E}\left|Y\right|^q < \infty$. Pick p, $q > 1$ such that $p^{-1} + q^{-1} = 1$. Then

$$\mathbb{E}\left|XY\right| \le \left(\mathbb{E}\mid X\mid^p\right)^{1/p}\left(\mathbb{E}\mid Y\mid^q\right)^{1/q}$$

When $p = q = 2$, this inequality is none other than the Cauchy-Schwartz inequality

$$\mathbb{E}(XY) \le \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}$$

Proof: If $X = 0$ or $Y = 0$ a.s., the inequality holds trivially. As such, suppose $x = \left(\mathbb{E}\mid X\mid^p\right)^{1/p} > 0$ and $y = \left(\mathbb{E}\mid Y\mid^q\right)^{1/q} > 0$. Suppose we are able to show that, for every point ω in the sample space,

$$\frac{\left|X(\omega)Y(\omega)\right|}{xy} \le \frac{1}{p}\frac{\left|X(\omega)\right|^p}{x^p} + \frac{1}{q}\frac{\left|Y(\omega)\right|^q}{y^p}$$

Then, taking expectations and rearranging, this becomes Holder's Inequality. To show this statement is true, we write $a = \left|X(\omega)\right| / x$ and $b = \left|Y(\omega)\right| / y$. The statement then becomes

$$ab \le \frac{a^p}{p} + \frac{a^q}{q}$$

With a and b non-negative. We can therefore write $a = e^{s/p}$ and $b = e^{t/q}$. The expression then follows trivially from the convexity of the exponential function:

$$\exp\left(\frac{s}{p} + \frac{t}{q}\right) \le \frac{e^s}{p} + \frac{e^t}{q}$$

We have therefore proved Holder's Inequality. ■

We will also require the following result

Lemma: If $p > 1$, then $\mathbb{E}\left|X\right|^p < \infty \Rightarrow \mathbb{E}\left|X\right|^q < \infty$ for $1 \le q \le p$.

Proof:

$$\begin{aligned}\mathbb{E}\left|X\right|^q &= \mathbb{E}\left(\left|X\right|^q;\left|X\right|<1\right)+\mathbb{E}\left(\left|X\right|^q;\left|X\right|\geq 1\right)\\ &\leq 1+\mathbb{E}\left(\left|X\right|^p;\left|X\right|\geq 1\right)\\ &\leq \infty\end{aligned}$$

As required. ■

Now, fix $K<\infty$ and $n\geq 1$, and choose p and q such that $p^{-1}+q^{-1}=1$. We then use Holder's Inequality to obtain

$$\begin{aligned}\mathbb{E}\left(\left|X_n\right|;\left|X_n\right|>K\right) &= \mathbb{E}\left(\left|X_n\right|\mathbb{I}_{\{|X_n|>K\}}\right)\\ &\leq \left(\mathbb{E}\left|X_n\right|^p\right)^{1/p}\left(\mathbb{E}\left(\mathbb{I}_{\{|X_n|>K\}}\right)^q\right)^{1/q}\end{aligned}$$

Taking a supremum

$$\mathbb{E}\left(\left|X_n\right|;\left|X_n\right|>K\right)\leq \sup_n\left(\mathbb{E}\left|X_n\right|^p\right)^{1/p}\left(\mathbb{P}(X_n>K)\right)^{1/q}$$

By the assumption in the theorem, the first expression is finite. Let its value be C:

$$\mathbb{E}\left(\left|X_n\right|;\left|X_n\right|>K\right)\leq C\left(\mathbb{P}(X_n>K)\right)^{1/q}$$

Using Markov's Inequality

$$\begin{aligned}\mathbb{E}\left(\left|X_n\right|;\left|X_n\right|>K\right) &\leq \left(\frac{\mathbb{E}\left|X_n\right|}{K}\right)^{1/q}\\ &\leq \left(\sup_n\frac{\mathbb{E}\left|X_n\right|}{K}\right)^{1/q}\end{aligned}$$

Now, it is clear that the first moment of X_n has finite expectation because we already know a *higher* moment has finite expectation:

$$\mathbb{E}\left(\left|X_n\right|;\left|X_n\right|>K\right)\leq \frac{CC'}{K^{1/q}}$$

The RHS now no longer depends on n, so we can write

$$\sup_n\mathbb{E}\left(\left|X_n\right|;\left|X_n\right|>K\right)\leq \frac{CC'}{K^{1/q}}$$

For any ε, provided we choose $K\geq (CC'/\varepsilon)^q$, our condition for uniform integrability is satisfied. ■

Proof (part 1 – alternative): We propose an alternative, somewhat simpler proof to part 1:

$$\mathbb{E}\left(\left|X_n\right|;\left|X_n\right|>a\right) \le \mathbb{E}\left(\left|X_n\right|\left\{\frac{\left|X_n\right|}{a}\right\}^{p-1};\left|X_n\right|>a\right)$$
$$=\frac{1}{a^{p-1}}\mathbb{E}\left(\left|X_n\right|;\left|X_n\right|>a\right)$$
$$=\frac{K}{a^{p-1}}$$

Setting $a \ge K / a^{p-1}$ gives the required result. ■

- ***Kolmogorov's 3-Series Theorem***
  - Let $\{X_n\}$ be a sequence of random variables, and let

$$S_n = \sum_{m=1}^{n} X_m$$

We now consider the question of *when* S_n converges. What condition is needed on the distribution of the X for this to happen? Clearly, if the X_n are IID, this will not be the case, since every additional variable in the sum add something to the sum – IID variables are either "dead at the start" or "never die out".

  - ***Theorem (Kolmogorov 3-series)***: Let $\{X_n\}$ be a sequence of *independent* random variables. Then $\sum_n X_n$ converges **if and only if** for some (and therefore every) $K > 0$, the following conditions hold:
    - $\sum_n \mathbb{P}\left(\left|X_n\right|>K\right)<\infty$ (clearly, for example, if the X_n are IID without bounded support, we'll never be able to satisfy this). This is a statement that the "mass in the tails" must decrease with n.
    - $\sum_n \mathbb{E}\left(X_n;\left|X_n\right|\le K\right)<\infty$
    - $\sum_n \mathbb{V}\text{ar}\left(X_n;\left|X_n\right|\le K\right)<\infty$

Note that there are no issues of *existence* of the expected value and the variance in the last two points because these are taken over a *finite* range – namely on $X_n \in (-K, K)$.

Proof: There is a simple proof of this result which relies on martingales. We will cover it later in this course.

- ***Proposition (2nd version of 3-series theorem)***: Let $\{X_n\}$ be independent random variables with $\mathbb{E}(X_n) = 0$ for all n. If $\sum_n \mathbb{V}\text{ar}(X_n) < \infty$, then $S_n = \sum_{m=1}^n X_m$ converges almost surely.

  Proof: Again, using martingale theory.

- ***Example***: Set $X_n = Y_n / n$, where the Y_n are IID exponential random variables with mean 1. Does the sum $S_n = \sum_n X_n$ converge? The deterministic series $1/n$ does *not* converge, and we wonder whether the exponential variables will be close enough to 0 often enough to "fix" that. In light of the result we derived in the previous lecture (that draws from an exponential will be arbitrarily large infinitely often) one might expect this series *not* to converge. Let us check the conditions formally.

  - *First condition*:

  $$\begin{aligned}\mathbb{P}\left(\left|X_n\right| > K\right) &= \mathbb{P}\left(\tfrac{Y_n}{n} > K\right) \\ &= \mathbb{P}\left(Y_n > nK\right) \\ &= e^{-nK}\end{aligned}$$

  This does indeed sum to a finite number

  $$\sum_n \mathbb{P}\left(\left|X_n\right| > K\right) < \infty$$

  The first condition is therefore met.

  - *Second condition*:

  $$\begin{aligned}\mathbb{E}\left(X_n ; \left|X_n\right| \le K\right) &= \mathbb{E}\left(\tfrac{Y_n}{n} ; Y_n \le nK\right) \\ &= \tfrac{1}{n}\mathbb{E}\left(Y_n ; Y_n \le nK\right)\end{aligned}$$

  But we note that

  $$\mathbb{E}\left(Y_1 ; Y_1 \le s\right) = \int_0^s y e^{-y} \ \mathrm{d}y > 0 \qquad \forall s > 0$$

  It is clear, therefore, that condition 2 is violated, because the sum of $1/n$ diverges. So, as we expected, S_n does not converge almost surely. □

- ***Example***: Set $X_n = Y_n/n$, where the Y_n are IID random variables with

  $$\mathbb{P}(Y_n = 1) = 1 - \mathbb{P}(Y_n = -1) = p$$

Once again, we wonder whether $S_n = \sum_{m=1}^{n} X_m$ converges. The intuition here is that the deterministic series $\sum_n (-1)^n / n$ *does* converge. Our version is this sum, but instead of deterministically flipping between positive and negative, we *randomly* switch between positive and negative. We wonder whether this will "spoil" the convergence. We fix $K > 0$ (say $K = 1$) and test the three conditions:

- *First condition*

  $$\mathbb{P}\left(\left|X_n\right| > 1\right) = 0$$

  So clearly, the sum does converge.

- *Second condition*

  $$\begin{aligned}\mathbb{E}\left(X_n;\left|X_n\right| \le 1\right) &= \frac{1}{n}\mathbb{E}\left(Y_n;\left|Y_n\right| \le n\right) \\ &= \frac{1}{n}\mathbb{E}(Y_n) \\ &= \frac{2p-1}{n}\end{aligned}$$

  We know, however, that $\sum \frac{k}{n}$ diverges, so the infinite sum of these expectations *only* converges if p = ½, in which case the numerator is 0. We therefore restrict our attention to that case.

- *Third condition* (*assuming* p = ½)

  $$\begin{aligned}\mathbb{V}\text{ar}\left(X_n;\left|X_n\right| \le 1\right) &= \frac{1}{n^2}\mathbb{V}\text{ar}\left(Y_n \mid \left|Y_n\right| \le n\right) \\ &= \frac{1}{n^2}\mathbb{V}\text{ar}(Y_n) \\ &= \frac{\mathbb{E}(Y^2) - 0}{n^2}\end{aligned}$$

  The infinite sum of the variances is therefore finite.

As such, our sum *does* converge. □

- ***Example***: Set $X_n = Y_n / \sqrt{n}$, where Y_n are IID random variables with probability ½ of being 1 and probably ½ of being –1. Again, the deterministic sum $\sum_n (-1)^n / \sqrt{n}$ converges, and we wonder whether adding randomness will make a difference. Since, in this case, $\mathbb{E}(X_n) = 0$, we can simply apply the second form of the 3-series theorem:

$$\begin{aligned}\mathbb{V}\text{ar}\left(X_n;\left|X_n\right|\le 1\right) &= \frac{1}{n}\mathbb{V}\text{ar}\left(X_n;\left|Y_n\right|\le n\right)\\ &= \frac{\mathbb{E}(X^2)}{n}\end{aligned}$$

The infinite sum of these variances clearly diverges, and so S_n does *not* converge. It is interesting to note how sensitive this result is to scaling – in the $1/n$ case, the convergence in the deterministic case carries over to the random case, whereas in the $1/\sqrt{n}$ case, it does not. □

- ***The Strong Law of Large numbers***
  - ***Theorem (SLLN)***: Let $\left\{X_n\right\}$ be IID with $\mathbb{E}\left|X_1\right|<\infty$, then

$$\frac{1}{n}\sum_{k=1}^n X_k \to \mathbb{E}(X_1) \qquad \text{a.s.}$$

Note: there is a simpler proof of this theorem that assumes the fourth moment of X_1 is bounded. We prove the more general theorem that does not require this stipulation.

Proof: Let us fix n.

- ***Step 1 – Truncation***: Let

$$Y_n = \begin{cases} X_n & \left|X_n\right|\le n \\ 0 & \text{otherwise}\end{cases}$$

In a sense, $\left\{Y_n\right\}$ is a "truncated" form of $\left\{X_n\right\}$. We will first prove the result for the truncated form, and later show that the result then carries over to the non-truncated form.

- ***Step 2***: Write Y_n as

$$Y_n = \overbrace{\left[Y_n - \mathbb{E}(Y_n)\right]}^{w_n} + \mathbb{E}(Y_n)$$

Now, note that since Y_n only contains values of $\left|X_n\right|$ that are less than or equal to n

$$\begin{aligned}\mathbb{E}(Y_n) &= \mathbb{E}\left(X_n;\left|X_n\right|\le n\right)\\ &= \mathbb{E}\left(X_n\mathbb{I}_{\{|X_n|\le n\}}\right)\end{aligned}$$

We will now attempt to evaluate this expectation using dominated convergence. Note that

1. $\left|X_1 \mathbb{I}_{\{|X_1| \leq n\}}\right| \leq |X_1|$
2. Since $|X_1|$ has finite mean, it is the case that

$$X_1 \mathbb{I}_{\{|X_1| \leq n\}} \to X_1 \mathbb{I}_{\{|X_1| < \infty\}} = X_1 \qquad \text{a.s.}$$

(1) justifies using dominated convergence, and (2) gives us the limit. We get

$$\mathbb{E}\left(X_1 \mathbb{I}_{\{|X_1| \leq n\}}\right) \to \mathbb{E}(X_1)$$
$$\mathbb{E}(Y_n) \to \mathbb{E}(X_1)$$

- ***Step 3***: We have shown that as $n \to \infty$, Y_n can be written as $Y_n = w_n + \mathbb{E}(X_1)$, or in other words that

$$\frac{1}{n}\sum_{k=1}^n Y_k \to \frac{1}{n}\sum_{k=1}^n w_k + \mathbb{E}(X_1)$$

The obvious next step is to show that the first term above tends to 0 almost surely. To do this, we will need two lemmas.

Lemma (Cesaro Sum Property): Let $\{a_n\}$ be a real valued sequence with $a_n \to a_\infty$. Then

$$\frac{1}{n}\sum_{k=1}^n a_k \to a_\infty \qquad \text{as } n \to \infty$$

Proof: Let $\varepsilon > 0$, and choose an N such that

$$a_k > a_\infty - \varepsilon \text{ whenever } k \geq N$$

Then

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{k=1}^n a_k \geq \liminf_n \left\{\frac{1}{n}\sum_{k=1}^N a_k + \frac{n-N}{n}(a_\infty - \varepsilon)\right\}$$
$$\geq 0 + a_\infty - \varepsilon$$

By a similar argument, $\limsup \leq a_\infty$. The result follows. ■

Lemma (Kronecker's Lemma): Let $\{a_n\}$ be a real valued sequence. Then

$$\left(\sum_{k=1}^{\infty}\frac{a_k}{k}\text{ converges}\right)\Rightarrow\left(\frac{1}{n}\sum_{k=1}^{n}a_k\to 0\right)$$

Proof: Let $u_n=\sum_{k\leq n}\frac{a_k}{k}$. By the assumption in our proof, $u_\infty=\lim_n u_n$ exists. Then

$$u_n-u_{n-1}=\frac{a_n}{n}$$

We then have

$$\sum_{k=1}^{n}a_k=\sum_{k=1}^{n}k(u_k-u_{k-1})=nu_n-\sum_{k=1}^{n}u_{k-1}$$

As such

$$\frac{1}{n}\sum_{k=1}^{n}a_k=u_n-\frac{1}{n}\sum_{k=1}^{n}u_{k-1}$$

The first term clearly tends to u_∞ as $n\to\infty$. By the Cesaro Sum Property, so does the second term. Thus

$$\frac{1}{n}\sum_{k=1}^{n}a_k\to u_\infty-u_\infty=0$$

As required. ■

Using these two lemmas, our (rough) strategy will be as follows:

- ***Step 3a***: Show that $\sum_{k=1}^{n}\frac{w_k}{k}$ converges almost surely, using the Kolmogorov 3-series theorem.
- ***Step 3b***: Use Kronecker's Lemma to deduce that for every outcome in the sample space for which $\sum_{k=1}^{n}\frac{w_k}{k}$ converges, $\frac{1}{n}\sum_{k=1}^{n}w_k$ converges to 0.
- ***Step 3c***: Note that since $\sum_{k=1}^{n}\frac{w_k}{k}$ converges almost surely, $\frac{1}{n}\sum_{k=1}^{n}w_k$ converges to 0 almost surely.

Step 3a: Note that since $w_k=Y_k-\mathbb{E}(Y_k)$, we can conclude that $\mathbb{E}(w_k)=0$. Furthermore, the w_k are independent of each other (because they are build up from the Y_k, which are built up from the X_k, which are independent). We can therefore use the second form of the Kolmogorov 3-series Theorem to show that $\sum_{k=1}^{n}\frac{w_k}{k}$ converges almost surely. (Note that some steps in the exposition below require some results about moments of random variable which we will prove in a homework – these are indicated by a *).

$$
\begin{aligned}
\mathbb{V}\mathrm{ar}\left(\tfrac{w_k}{k}\right) &= \frac{1}{k^2}\mathbb{V}\mathrm{ar}\left(Y_k - \mathbb{E}(Y_k)\right) \\
&= \frac{1}{k^2}\mathbb{V}\mathrm{ar}\left(Y_k\right) \\
&= \frac{\mathbb{E}(Y_k^2) - \mathbb{E}(Y_k)^2}{k^2} \\
&\le \frac{\mathbb{E}(Y_k^2)}{k^2} \\
{}^* &= \frac{1}{k^2}\int_0^\infty 2y\mathbb{P}\left(\left|Y_k\right| > y\right)\,\mathrm{d}y \\
&= \frac{1}{k^2}\int_0^k 2y\mathbb{P}\left(\left|X_k\right| > y\right)\mathrm{d}y \\
&= \frac{1}{k^2}\int_0^k 2y\mathbb{P}\left(\left|X_1\right| > y\right)\mathrm{d}y
\end{aligned}
$$

As such

$$
\sum_{k=1}^\infty \mathbb{V}\mathrm{ar}\left(\tfrac{w_k}{k}\right) \le \sum_{k=1}^\infty \frac{1}{k^2}\int_0^k 2y\mathbb{P}\left(\left|X_1\right| > y\right)\,\mathrm{d}y
$$

By Frobini's Lemma, we can interchange the summation and integration, since the integrands are all positive

$$
\begin{aligned}
\sum_{k=1}^\infty \mathbb{V}\mathrm{ar}\left(\tfrac{w_k}{k}\right) &\le \int_0^\infty \sum_{k=1}^\infty \frac{1}{k^2}\mathbb{I}_{\{y\le k\}} 2y\mathbb{P}\left(\left|X_1\right| > y\right)\mathrm{d}y \\
&\le \int_0^\infty 2y\mathbb{P}\left(\left|X_1\right| > y\right)\sum_{k=\lfloor y\rfloor}^\infty \frac{1}{k^2}\,\mathrm{d}y
\end{aligned}
$$

Note, however, that the sum in this expression above can – in theory – be bounded by some integral (for example, $1/(k-1)^2$ works well). Thus, we can write

$$
\begin{aligned}
\sum_{k=1}^\infty \mathbb{V}\mathrm{ar}\left(\tfrac{w_k}{k}\right) &\le \int_0^\infty \frac{C_1}{C_2 + y} 2y\mathbb{P}\left(\left|X_1\right| > y\right) \\
&\le C_3\int_0^\infty \mathbb{P}\left(\left|X_1\right| > y\right) \\
{}^* &= C_3\mathbb{E}\left|X_1\right| \\
&< \infty
\end{aligned}
$$

So the sum does indeed converge.

- ***Step 3b & 3c***: Consider the following two events

$$A = \left\{ \omega : \sum_{k=1}^{\infty} \tfrac{w_k(\omega)}{k} < \infty \right\}$$
$$B = \left\{ \omega : \frac{1}{n} \sum_{k=1}^{\infty} w_k(\omega) \to 0 \right\}$$

Now, by Kronecker's Lemma, $A \subseteq B$. But by Step 3a, $\mathbb{P}(A) = 1$, since A occurs almost surely. Thus, $\mathbb{P}(B) = 1$ and

$$\frac{1}{n} \sum_{k=1}^{\infty} w_k \to 0 \qquad \text{a.s.}$$

Thus, from Step 2,

$$\frac{1}{n} \sum_{k=1}^{n} Y_k \to \mathbb{E}(X_1) \qquad \text{a.s.}$$

- ***Step 4***: We have now proved the SLLN for the truncated series $\left\{Y_n\right\}$. All that remains to be shown is that the result also applies to the full series $\left\{X_n\right\}$. Now, consider that

  $$\mathbb{P}\left(X_n \neq Y_n\right) = \mathbb{P}\left(\left|X_n\right| > n\right)$$

  But note that

  $$\sum_n \mathbb{P}\left(\left|X_n\right| > n\right) = \sum_n \mathbb{P}\left(\left|X_1\right| > n\right) \leq \mathbb{E}(X_1) < \infty$$

  So by the first BC Lemma, for large enough n $\left|X_n\right| \leq n$ almost surely. Thus, for large enough n, $X_n = Y_n$ almost surely. Now, consider that, by the triangle inequality and the Cesaro sum property

  $$\begin{aligned}\left| \tfrac{1}{n} \sum_{k=1}^{n} X_k(\omega) - \tfrac{1}{n} \sum_{k=1}^{n} Y_k(\omega) \right| &\leq \tfrac{1}{n} \sum_{k=1}^{n} \left| X_k(\omega) - Y_k(\omega) \right| \\ &\to \lim_n \left| X_n(\omega) - Y_n(\omega) \right|\end{aligned}$$

  But we have already shown that $\lim_n \left| X_n(\omega) - Y_n(\omega) \right| = 0$ almost surely. As such

  $$\left| \tfrac{1}{n} \sum_{k=1}^{n} X_k(\omega) - \tfrac{1}{n} \sum_{k=1}^{n} Y_k(\omega) \right| \to \lim_n \left| X_n(\omega) - Y_n(\omega) \right| \qquad \text{a.s.}$$

  Thus, the SLLN holds for the original sequence X_n. ■

- ***Example***: Let $\left\{X_n\right\}$ be an IID sequence of random variables with $\mathbb{E}(X_n) = 0$ and $\mathbb{E}\left(X_n^2\right) = \sigma^2 < \infty$. Let $S_n = \sum_{k=1}^{n} X_n$. We know, from the SLLN, that

  $$\frac{S_n}{n} \to 0 \qquad \text{a.s}$$

We might wonder, however, whether this still holds true if, instead of dividing by n, we divide by some quantity a_n, where $a_n \nearrow \infty$, but possibly slower than n. We might wonder what the *smallest* such a_n is that still makes the sum converge.

Let us consider, for example, $a_n = \sqrt{n}\left(\log n\right)^{\frac{1}{2}+\varepsilon}$. log n grows extremely slowly, so this is really very close to a $\sqrt{n}$ growth. Consider that

$$\mathbb{V}\text{ar}\left(\frac{X_n}{a_n}\right) = \frac{\sigma^2}{\sqrt{n}\left(\log n\right)^{\frac{1}{2}+\varepsilon}}$$

And this implies that

$$\sum_n \mathbb{V}\text{ar}\left(\tfrac{X_n}{a_n}\right) < \infty$$

Thus, by the Kolmogorov 3-series theorem, $\sum_n \frac{X_n}{a_n}$ converges almost surely, and by Kronecker's Lemma, this means that $S_n / a_n \to 0$. Thus, we see that even with such a small-growing a_n, an almost-sure result still holds.

However, it is interesting to note that if we take $a_n = \sqrt{n}$ (slightly smaller), then *no almost sure result* holds anymore. In fact, we will show in the next lecture that the best we can say is $S_n / \sqrt{n} \to \sigma N(0,1)$. These arguments are clearly extremely sensitive to scaling. □

Lecture 3 – *2nd February 2011*

- ***Weak convergence & the Central Limit Theorem***
  - ***Definition*** (***Weak convergence***): A sequence of random variables $\{X_n\}$ is said to converge weakly to a random variable X, denoted $X_n \Rightarrow X$ as $n \to \infty$, if and only if

    $$\mathbb{E}\left[f(X_n)\right] \to \mathbb{E}\left[f(X)\right]$$

    For all bounded continuous functions f.

    Comments:

    - This is clearly a very difficult definition to use, because there are "too many" bounded continuous function. We will see later, however, that it suffices to consider a smaller set of functions that is a "basis" for these functions.

- If $X_n \to X$ almost surely, then $f(X_n) \to f(X)$, because f is continuous, and $\mathbb{E}\left[f(X_n)\right] \to \mathbb{E}\left[f(X)\right]$, by bounded convergence (since the functions are bounded). Thus, almost sure convergence implies weak convergence.
- Each member of $\{X_n\}$ as well as X need not even live on the same probability space. Indeed, the theorem only mentions convergence of expectations, which are effectively integrals

$$\int f(x)\ \mathrm{d}F_n(x) \to \int f(x)\ \mathrm{d}F(x)$$

Nothing prevents the integrals from being taken over different spaces. Note that this is radically different to the context of almost sure convergence, which *fixes* an ω in the probability space and requires convergence. This is not possible here, because the variables might not even live in the same probability space.

- ***Definition (Weak convergence)***: A sequence of *real valued* random variables $\{X_n\}$ converges weakly to X if and only if

$$F_n(x) \to F(x)$$

for all continuity points of F, where $X_n \sim F_n$ and $X \sim F$. (This definition makes it even clearly that the variables can live in different spaces).

- ***Example***: Take $X_1, X_2, \cdots$ IID, with $\mathbb{E}(X_1) = \mu$ and $\mathbb{V}\mathrm{ar}(X_1) = \sigma^2$, then

$$\frac{\sum_{i=1}^{\infty}(X_i - \mu)}{\sqrt{n}} \Rightarrow \sigma N(0,1)$$

This is none other than the central limit theorem. (Note the slight abuse of notation which will continue throughout the course – when we say $\Rightarrow F$, we mean $\Rightarrow X$ where $X \sim F$). We will prove this late in this lecture. □

- ***Example (Birthday Problem)***: Let $X_1, X_2, \cdots$ be IID uniformly distributed on $\{1, \cdots, n\}$ and independent of each other (if the X_n were birthdays, we would have $n = 365$). Now, fix $k < n$ (the number of people amongst whom we are looking for a match), and let

$$Y_n = \min\left\{i : X_i = X_j \text{ for some } j < i\right\}$$

This is effectively the smallest i for which X_i matches an earlier X. We are now interested in the quantity

$$\begin{aligned}\mathbb{P}\left(\text{No match in group of } k \text{ people}\right) &= \mathbb{P}\left(Y_n > k\right) \\ &= 1 \cdot \frac{n-1}{n} \cdot \frac{n-2}{n} \cdot \dots \cdot \frac{n-k+1}{n}\end{aligned}$$

We *could* plug in numbers of n and k into the expression above and calculate the product. This is somewhat inconvenient – let's try to use weak convergence arguments to get an approximation to this instead. Let us fix $x > 0$

$$\begin{aligned}\mathbb{P}\left(\frac{Y_n}{\sqrt{n}} > x\right) &= \mathbb{P}\left(Y_n > x\sqrt{n}\right) \\ &= \prod_{i=2}^{\lfloor x\sqrt{n}\rfloor}\left(\tfrac{n-i+1}{n}\right) \\ &= \prod_{i=2}^{\lfloor x\sqrt{n}\rfloor}\left(1 - \tfrac{i-1}{n}\right)\end{aligned}$$

Let us take logarithms of both sides

$$\log \mathbb{P}\left(\frac{Y_n}{\sqrt{n}} > x\right) = \sum_{i=2}^{\lfloor x\sqrt{n}\rfloor} \log\left(1 - \tfrac{i-1}{n}\right)$$

We now need a quick definition

> ***Definition***: $f(x) \sim g(x)$ as $x \to \infty$ if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$. Similarly, $a_n \sim b_n$ if $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$. Note that this does not necessarily imply either function/series converges.

Now, note that $\log(1-x) \sim -x$ as $x \to 0$. As such, $\log\left(1 - \frac{y}{n}\right) \sim -\frac{y}{n}$. Therefore

$$\begin{aligned}\log \mathbb{P}\left(\frac{Y_n}{\sqrt{n}} > x\right) &\sim -\sum_{i=2}^{\lfloor x\sqrt{n}\rfloor} \frac{i-1}{n} \\ &= -\frac{1}{n}\sum_{i=2}^{\lfloor x\sqrt{n}\rfloor}(i-1) \\ &\sim -\frac{1}{n}\frac{1}{2}x^2 n \\ &= -x^2/2\end{aligned}$$

As such

$$\mathbb{P}\left(\frac{Y_n}{\sqrt{n}} > x\right) \to \exp\left(-\tfrac{1}{2}x^2\right)$$

The RHS is clearly a distribution, because it takes value 1 at $x = 0$ and decreases thereafter, and we therefore have weak convergence by the second definition.

Before we comment on this result, let us quote a quick Lemma

Lemma: $\left(\frac{1}{x} - \frac{1}{x^3}\right)e^{-x^2/2} \leq \int_x^\infty e^{-y^2/2}\ \mathrm{d}y \leq \frac{1}{x}e^{-x^2/2}$

Proof: For the upper bound:

$$\int_x^\infty e^{-y^2/2}\ \mathrm{d}y \overbrace{\leq}^{\text{Integral over } y \geq x} \int_x^\infty \frac{y}{x} e^{-y^2/2}\ \mathrm{d}y = \frac{1}{x}e^{-x^2/2}$$

As required. For the lower bound:

$$\int_x^\infty e^{-y^2/2}\ \mathrm{d}y \geq \int_x^\infty \left(1 - \frac{3}{y^4}\right)e^{-y^2/2}\ \mathrm{d}y = \left(\frac{1}{x} - \frac{1}{x^3}\right)e^{-x^2/2}$$

(Which can be shown by differentiating the result). ■

Corollary: $\int_x^\infty e^{-y^2/2}\ \mathrm{d}y \sim \frac{1}{x}e^{-x^2/2}$ as $x \to \infty$. Making this more specific to a normally distributed random variable, suppose $X \sim N(0,\sigma^2)$, then

$$\begin{aligned}\mathbb{P}\left(X > x\right) &= \int_x^\infty \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{-y^2/2\sigma^2}\ \mathrm{d}y \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}\sigma \int_{x/\sigma}^\infty \exp^{-\psi^2/2}\ \mathrm{d}\psi \\ &\sim \frac{1}{\sqrt{2\pi}}\frac{\sigma}{x}e^{-x^2/2\sigma^2}\end{aligned}$$

It is interesting to note, therefore, how similar our result looks to the central limit theorem, despite the fact we did not center the original variables Y_n. The distribution we eventually obtained has a tail integral, which, by the corollary, looks similar to that for a random variable. (In fact, the random variable with this tail integral is a Rayleigh random variable). □

Remark: A full rigorous treatment of this example will be done in homework 1.

- ***Example***: Let $X_1, X_2, \cdots$ be IID $N(0,\sigma^2)$ random variables, and let $M_n = \max\left(X_1, \cdots, X_n\right)$. Consider the case $\sigma^2 = 1$, and let α_n be such that $\mathbb{P}\left(X_1 > \alpha_n\right) = \frac{1}{n}$ (we will show in homework 1 that $\alpha_n \sim \sqrt{2\log n}$). We would like to show that

$$\mathbb{P}\left\{\alpha_n(M_n - \alpha_n) \le x\right\} \to \exp\left\{-e^{-x}\right\} \qquad \forall x \in \mathbb{R}$$

We first note the RHS has no discontinuity points. Next, note that as $x \to \infty$, the RHS tends to 1, whereas as $x \to -\infty$, the RHS tends to 0. Finally, we note the RHS is monotone. Thus, the RHS is a valid cumulative distribution function. Let us now show convergence:

$$\begin{aligned}\mathbb{P}\left(M_n \le \tfrac{x}{\alpha_n} + \alpha_n\right) &= \mathbb{P}\left(X_1 \le \tfrac{x}{\alpha_n} + \alpha_n, \cdots, X_n \le \tfrac{x}{\alpha_n} + \alpha_n\right) \\ &= \left[\mathbb{P}\left(X_1 \le \tfrac{x}{\alpha_n} + \alpha_n\right)\right]^n \\ &= \left[1 - \mathbb{P}\left(X_1 > \tfrac{x}{\alpha_n} + \alpha_n\right)\right]^n\end{aligned}$$

We will now require a lemma

> ***Lemma***: Consider two (possibly non-real) sequences $\left\{a_n\right\}$ and $\left\{b_n\right\}$. If $a_n \to 0$, $b_n \to \infty$ and $a_n b_n \to c$, then $(1 + a_n)^{b_n} \to e^c$ as $n \to \infty$.

This is precisely the situation in this case, with $b_n = n \to \infty$ and $a_n = -\mathbb{P}\left(X_1 > \tfrac{x}{\alpha_n} + \alpha_n\right) \to 0$, since $\alpha_n \to \infty$. We must now consider whether $a_n b_n \to 0$. To do this, we will make extensive use of the Lemma in the previous example (namely that $\mathbb{P}\left(X_1 > x\right) = \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}$)

$$\begin{aligned}n \cdot \mathbb{P}\left(X_1 > \tfrac{x}{\alpha_n} + \alpha_n\right) &\sim n \cdot \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\frac{x}{\alpha_n} + \alpha_n} \cdot \exp\left\{-\tfrac{1}{2}\left(\tfrac{x}{\alpha_n} + \alpha_n\right)^2\right\} \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\frac{x}{\alpha_n} + \alpha_n} n \cdot \exp\left\{-\tfrac{x^2}{2\alpha_n^2} - x - \tfrac{\alpha_n^2}{2}\right\} \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\frac{x}{\alpha_n} + \alpha_n} n \cdot e^{-x} e^{-x^2/2\alpha_n^2} e^{-\alpha_n^2/2} \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\frac{x}{\alpha_n} + \alpha_n} n \cdot e^{-x} e^{-x^2/2\alpha_n^2} \alpha_n \sqrt{2\pi} \frac{1}{\alpha_n \sqrt{2\pi}} e^{-\alpha_n^2/2} \\ &\sim \frac{1}{\sqrt{2\pi}} \frac{1}{\frac{x}{\alpha_n} + \alpha_n} n \cdot e^{-x} e^{-x^2/2\alpha_n^2} \alpha_n \sqrt{2\pi} \overbrace{\mathbb{P}\left(X_1 > \alpha_n\right)}^{=1/n} \\ &= \frac{\alpha_n}{\frac{x}{\alpha_n} + \alpha_n} e^{-x} e^{-x^2/2\alpha_n^2} \\ &\to e^{-x}\end{aligned}$$

As such, by the Lemma

$$\mathbb{P}\left(M_n \le \tfrac{x}{\alpha_n} + \alpha_n\right) = \left[1 - \mathbb{P}\left(X_1 > \tfrac{x}{\alpha_n} + \alpha_n\right)\right]^n$$
$$\to e^{-e^{-x}}$$

As required. □

- ***Definition (Tightness)***: A sequence of random variables $\{X_n\}$ with corresponding distribution functions $\{F_n\}$ is said to be *tight* if $\forall \varepsilon > 0$, $\exists K(\varepsilon) < \infty$ such that

$$\sup_n \mathbb{P}\left\{\left|X_n\right| > K(\varepsilon)\right\} \le \varepsilon$$

Equivalently

$$\sup_n \left[1 - F_n\left(K(\varepsilon)\right)\right] \le \varepsilon$$

Comment: A sufficient condition for tightness if $\sup_n \mathbb{E}\left|X_n\right| < \infty$. To see why, with $K > 0$, and consider that by Markov's Inequality

$$\mathbb{P}\left(\left|X_n\right| > K\right) \le \frac{\mathbb{E}\left|X_n\right|}{K} \le \frac{\overbrace{\sup_n \mathbb{E}\left|X_n\right|}^{<\infty}}{K}$$

- ***Proposition***: Let $\{X_n\}$ be a sequence of random variables
  - If $X_n \Rightarrow X$, then $\{X_n\}$ is tight
  - If $\{X_n\}$ is tight, then there exists a subsequence $\{n_k\}$ such that $\{X_{n_k}\}$ converges weakly.

This is somewhat isomorphic to the concept of compactness in real analysis.

- ***Definition (Characteristic Function)***: A characteristic function (CF) of a random variable X is given by

$$\varphi_X(\theta) = \mathbb{E}\left[\exp(i\theta X)\right] \qquad \theta \in \mathbb{R}$$

Remarks

  - $\varphi_X(0) = 1$
  - $\left|\varphi_X(\theta)\right| \le 1$, because by Jensen's inequality, $\left|\varphi_X(\theta)\right| \le \left\{\mathbb{E}\left|\exp(i\theta X)\right|\right\} = 1$
  - $\varphi_X^{(n)}(0) = i^n \mathbb{E}(X^n)$, provided $\mathbb{E}\left|X\right|^n < \infty$.
  - If X and Y are independent, $\varphi_{X+Y}(\theta) = \varphi_X(\theta)\varphi_Y(\theta)$, though the converse is false.

- ***Examples***: If $X \sim N(\mu, \sigma^2)$, then

$$\varphi_X(\theta) = \exp\left(i\mu\theta - \tfrac{1}{2}\sigma^2\theta^2\right)$$

Note also that if $Y = \mu + \sigma X$, then

$$\varphi_Y(\theta) = \mathbb{E}\left(e^{i\theta[\mu+\sigma X]}\right) = e^{i\theta\mu} + \varphi_X(\sigma\theta)$$

This is convenient in relating *all* normally distributed random variables to the standard normal. □

- ***Proposition (Levy characterization theorem)***: Let $\{X_n\}$ be a sequence of random variables with distribution functions $\{F_n\}$. If
  - $\varphi_{X_n}(\theta) \to \varphi(\theta)$ for all $\theta \in \mathbb{R}$
  - $\varphi(\theta)$ is continuous at 0.

  Then

$$X_n \Rightarrow X$$

  and X has characteristic function $\varphi(\theta)$.

  Comments:
  - If $X_n \Rightarrow X$, then the two conditions in the proposition trivially hold – this is really an "if and only if" statement.
  - One might wonder why the second condition in the theorem is needed. To see why, consider, for example, $X_n \sim N(0, n)$. We then have $\varphi_{X_n}(\theta) = e^{-\theta^2 n/2}$, and as n grows large, this converges to

$$\varphi_{X_n}(\theta) \to \begin{cases} 0 & \text{if } \theta \neq 0 \\ 1 & \text{if } \theta = 0 \end{cases}$$

    This does not satisfy our continuity requirements at 0. To see why, consider that with $K > 0$,

$$\mathbb{P}\left(\left|X_n\right| > K\right) = \mathbb{P}\left(N(0,1) > \frac{K}{\sqrt{n}}\right) \to \frac{1}{2}$$

    In other words, the distribution in question is not tight – tail probabilities grown with n. The continuity condition, therefore, can be thought of as a tightness condition.

- ***The Central Limit Theorem***
  - ***Theorem***: Let $X_1, X_2, \cdots$ be IID with mean $\mathbb{E}(X_1) = \mu$ and variance $\sigma^2 = \mathbb{V}\text{ar}(X_1)$. Then

$$\frac{\sum_{i=1}^{n}(X_i - \mu)}{\sqrt{n}} \Rightarrow \sigma N(0,1)$$

Proof: It suffices to prove this for the case of $\mu = 0$. Let $S_n = \sum_{i=1}^{n} X_i$, and consider its characteristic function

$$\begin{aligned}
\varphi_{S_n/\sqrt{n}}(\theta) &= \mathbb{E}\left[\exp\left(i\theta S_n / \sqrt{n}\right)\right] \\
&= \mathbb{E}\left[\exp\left(\tfrac{i\theta}{\sqrt{n}}\sum_{j=1}^{n} X_j\right)\right] \\
&= \mathbb{E}\left[\prod_{j=1}^{n}\exp\left(\tfrac{i\theta}{\sqrt{n}}X_j\right)\right] \\
&= \prod_{j=1}^{n}\mathbb{E}\left[\exp\left(\tfrac{i\theta}{\sqrt{n}}X_j\right)\right] \\
&= \left[\varphi_X\left(\tfrac{\theta}{\sqrt{n}}\right)\right]^n
\end{aligned}$$

Now, let us take a Taylor Expansion

$$\begin{aligned}
\varphi_X\left(\tfrac{\theta}{\sqrt{n}}\right) &= \varphi_X(0) + \frac{\theta}{\sqrt{n}}\varphi_X'(0) + \frac{\theta^2}{2n}\varphi_X''(0) + R_n \\
&= \varphi_X(0) + \frac{\theta}{\sqrt{n}}\mathbb{E}(X) + \frac{\theta^2}{2n}i^2\mathbb{E}(X^2) + R_n \\
&= 1 - \frac{\theta^2\sigma^2}{2n} + R_n
\end{aligned}$$

Now, consider that

$$\begin{aligned}
\varphi_{S_n/\sqrt{n}}(\theta) &= \left(\varphi_X\left(\tfrac{\theta}{\sqrt{n}}\right)\right)^n \\
&= \left(1 - \frac{\theta^2\sigma^2}{2n} + R_n\right)^n \\
&= \left(1 + a_n\right)^{b_n}
\end{aligned}$$

Now, suppose we can show that $a_n b_n = n\left[\frac{\theta^2\sigma^2}{2n} + R_n\right] = \frac{\theta^2\sigma^2}{2} + nR_n \to \frac{\theta^2\sigma^2}{2}$ (or in other words, that $nR_n \to 0$) as $n \to \infty$, then we would have

$$\begin{aligned}
\varphi_{S_n/\sqrt{n}}(\theta) &\to \exp\left(-\frac{\theta^2\sigma^2}{2}\right) \\
&= \varphi(\theta)
\end{aligned}$$

This is the characteristic function of $N(0,\sigma^2)$, and it is continuous at $\theta = 0$. Thus, all the conditions of Levy's Theorem hold, which proves the central limit theorem.

All we now need to do is to prove that $nR_n \to 0$ as $n \to \infty$. To do this, first note the following standard result from deterministic analysis proof?!?

$$\left| e^{ix} - \sum_{m=0}^{n} \frac{(ix)^n}{m!} \right| \le \left(\frac{|x|^{n+1}}{(n+1)!} \wedge \frac{2|x|^n}{n!} \right)$$

Here, $a \wedge b = \min(a,b)$. Let us apply expectations

$$\mathbb{E}\left| e^{ix} - \sum_{m=0}^{n} \frac{(ix)^n}{m!} \right| \le \mathbb{E}\left(\frac{|x|^{n+1}}{(n+1)!} \wedge \frac{2|x|^n}{n!} \right)$$

Now, use Jensen's Inequality on the LHS to obtain

$$\left| \mathbb{E}\left(e^{i\theta X}\right) - \mathbb{E}\left(\sum_{m=0}^{2} \frac{(iX\theta)^m}{m!} \right) \right| \le \mathbb{E}\left(\frac{|X\theta|^2}{3!} \wedge \frac{2|X\theta|^2}{2} \right) = \mathbb{E}\left[f(X,\theta) \right]$$

Now, observe that if $\theta \to 0$, $f(X,\theta) \to 0$ almost surely, since $\mathbb{E}|X|$ is bounded. Similarly, note that $f(X,\theta)$ is bounded by $|X|^2 \theta^2$, and since $\mathbb{E}|X|^2 = \sigma^2 < \infty$, we can say that $f(X,\theta) \le Y$ with $\mathbb{E}|Y| < \infty$. We can therefore apply the dominated convergence theorem to conclude that

$$\mathbb{E}\left[f(X,\theta) \right] \to 0 \qquad \text{as } \theta \to 0$$

More details?

But we have

$$\begin{aligned} R_n &= \mathbb{E}\left(f(X, \tfrac{\theta}{\sqrt{n}}) \right) \\ &= n\mathbb{E}\left\{ \frac{\left| X \frac{\theta}{\sqrt{n}} \right|^3}{3!} \wedge |X|^2 \frac{\theta^2}{n} \right\} \\ &= \mathbb{E}\left\{ \frac{|X|^3 \frac{\theta^3}{\sqrt{n}}}{3!} \wedge |X|^2 \frac{\theta^2}{n} \right\} \end{aligned}$$

Inside goes to 0 as n -> infinity for all theta, and is dominated by |X|^2 theta^2 which has finite expectation, hence by dominated convergence it all goes to 0. We kind of need thrd moment – but clever combination here using second moments as well to be able to use interchange.

- ***Odds & Ends***
  - ***Proposition (The Skorohod Representation)***: Let $\{X_n\}$ be a sequence of random variables such that $X_n \Rightarrow X$. Then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ supporting a sequence $\{X_n'\}$ and X' such that

$$X_n' \overset{d}{=} X_n \text{ and } X' \overset{d}{=} X$$

And $X_n' \to X'$ almost surely.

Proof: Let F_n be the distribution of X_n and F be the distribution of X. We know that

$$F_n(x) \to F(x) \qquad (\star)$$

at all continuity points x of F. Now, define

$$F^{-1}(p) = \inf\left\{x : F(x) \geq p\right\}$$

(This generalized inverse deals with cases in which there are discontinuities in the distribution). Note that $(\star)$ implies that

$$F_n^{-1}(p) \to F^{-1}(p) \qquad (\star\star)$$

How do we know? for all continuity points of F (and recall that by the definition of distribution functions, F only has a finite number of discontinuity points).

Now, take $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{F}, \text{uniform distribution})$. Let $U \sim U[0,1]$. Let

$$X_n' = F_n^{-1}(U) \qquad X' = F^{-1}(U)$$

Note that

$$\begin{aligned}\mathbb{P}\left(X_n' \leq x\right) &= \mathbb{P}\left(F_n'(U) \leq x\right) \\ &= \mathbb{P}\left(U \leq F_n(x)\right) \\ &= F_n(x)\end{aligned}$$

Similarly, $\mathbb{P}\left(X' \leq x\right) = F(x)$. As such, $X' \overset{d}{=} X$ and $X_n' \overset{d}{=} X_n$, as required. Note also, however, that by $(\star\star)$, $X_n' \to X'$ for all but continuity points in Ω. Since there are a countable number of such points, the set of such points has measure 0. Thus, $X_n' \to X'$ almost surely.

- ***Proposition (Continuous Mapping Theorem – CMT)***: Let $\{X_n\}$ be such that $X_n \Rightarrow X$. Let f be a function with $\mathbb{P}\left(X \in D_f\right) = 0$ with D_f being the set of discontinuity points of f. Then $f(X_n) \Rightarrow f(X)$
- ***Proposition (Converging Together Lemma*** – CTL): If $\{X_n\}$ is a sequence such that $X_n \Rightarrow X$ and $\{Y_n\}$ is a sequence such that $Y_n \Rightarrow a$ (where a is a deterministic constant), then
  - $X_n + Y_n \Rightarrow X + a$
  - $X_n Y_n \Rightarrow Xa$
  - $\frac{X_n}{Y_n} \Rightarrow \frac{1}{a} X$ provided $a \neq 0$.

LECTURE 4 – *9th February 2011*

- ***Introduction to Large Deviations***
  - Consider a sequence of IID random variables, $X_1, X_2, \cdots \sim N(\mu, \sigma^2)$. Let $S_n = \sum_{i=1}^n X_i$. Fix $a > \mu$. We are interested in the value

$$\mathbb{P}\left(S_n > na\right)$$

In the case of the Gaussian distribution, we can use the bounds derived earlier in these notes

$$\begin{aligned}\mathbb{P}\left(S_n > na\right) &= \mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} > \frac{n(a-\mu)}{\sigma\sqrt{n}}\right) \\ &= \mathbb{P}\left(Z > \sqrt{n}\,\frac{a-\mu}{\sigma}\right) \\ &\leq \frac{1}{\sqrt{2\pi}}\frac{\sigma}{\sqrt{n}(a-\mu)}\exp\left(-n\frac{(a-\mu)^2}{2\sigma^2}\right)\end{aligned}$$

And

$$\mathbb{P}\left(S_n > na\right) \geq \frac{1}{\sqrt{2\pi}}\left(\left(\tfrac{\sqrt{n}(a-\mu)}{\sigma}\right)^{-1} - \left(\tfrac{\sqrt{n}(a-\mu)}{\sigma}\right)^{-3}\right)\exp\left(-\frac{n(a-\mu)^2}{2\sigma^2}\right)$$

Taking logarithms, we find that (c_1, c_2 and c_3 are constants)

$$\frac{c_2}{n}+\frac{c_3\log n}{n}-\frac{(a-\mu)^2}{2\sigma^2}\le\frac{1}{n}\log\mathbb{P}\left(S_n>na\right)\le\frac{c_1}{n}-\frac{\frac{1}{2}\log n}{n}-\frac{(a-\mu)^2}{2\sigma^2}$$

Now, this means that

$$\limsup_n\frac{1}{n}\log\mathbb{P}\left(S_n>na\right)\le-\frac{(a-\mu)^2}{2\sigma^2}$$
$$\liminf_n\frac{1}{n}\log\mathbb{P}\left(S_n>na\right)\ge-\frac{(a-\mu)^2}{2\sigma^2}$$

As such, we find that

$$\frac{1}{n}\log\mathbb{P}\left(S_n>na\right)\to-\frac{(a-\mu)^2}{2\sigma^2}$$

This result is the *meta* result of large deviations theory. It is easy to derive in the case of a Gaussian variable – we seek to develop this result in more generality.

- Let us develop some insight into the exact concept of "*large* deviations". The central limit theorem would tell us that

$$S_n\approx n\mu+\sigma Z\sqrt{n}\qquad\qquad Z\sim N(0,1)$$

A "natural" question, therefore, would be to consider

$$\mathbb{P}(S_n>n\mu+\sigma K\sqrt{n})$$

The theory of large deviations goes even further – since $a>\mu$, we can write

$$\mathbb{P}(S_n>na)=\mathbb{P}(S_n>n\mu+\sigma Kn)$$

In other words, the central limit theorem deals with "normal" deviations, of order $\sqrt{n}$ from the mean. The theory of large deviations refers to larger deviations, of order n. We will now see why the central limit theorem is powerless to deal with these large deviations.

- Consider a sequence of IID random variables, $X_1,X_2,\cdots$, with $\mathbb{E}(X_1)=\mu$ and $\mathbb{V}\text{ar}(X_1)=\sigma^2<\infty$ As ever, let $S_n=\sum_{i=1}^n X_i$. Fix $a>\mu$. We might be tempted to naively apply the Central Limit Theorem and write

$$\begin{aligned}\mathbb{P}\left(S_n>na\right)&=\mathbb{P}\left(\frac{S_n-n\mu}{\sigma\sqrt{n}}>\frac{n(a-\mu)}{\sigma\sqrt{n}}\right)\\&\to\mathbb{P}\left(Z>\sqrt{n}\frac{a-\mu}{\sigma}\right)\\&\to 0\qquad\text{as } n\to\infty\end{aligned}$$

Though this statement is correct, it isn't particularly informative – it's quite obvious that the probability in question tends to 0! This happens because of the factor of $\sqrt{n}$ on the RHS of the inequality. The central limit theorem is clearly not equipped to deal with large deviations.

- ***Theorem***: Let $X_1, X_2, \cdots$ be IID with mean μ and such that $M(\theta) = \mathbb{E}(e^{\theta X_1})$ exists for all $\theta \in \mathbb{R}$, and assume $\mathbb{P}(X_1 = \mu) < 1$. Then for any $a > \mu$,

$$\frac{1}{n} \log \mathbb{P}\left(S_n > na\right) \to -I(a)$$

Where

$$I(a) = \sup_{\theta \in \mathbb{R}} \left[\theta a - \log M(\theta)\right]$$
$$I(a) = \sup_{\theta \in \mathbb{R}} \left[\theta a - \varphi(\theta)\right]$$

(We denote $\varphi(\theta) = \log M(\theta)$). This is the *convex conjugate* of φ.

Remark: The same conclusions hold if $M(\theta)$ exists in a neighborhood of the origin (plus some mild technicalities).

- ***Example***: Let us apply this theorem to Gaussian random variables, for which $M(\theta) = \exp\left(\theta\mu + \frac{1}{2}\theta^2\sigma^2\right)$. In this case,

$$I(a) = \sup_\theta \left\{\theta a - \theta\mu - \tfrac{1}{2}\theta^2\sigma^2\right\}$$

This is simply a quadratic, with argmax $\theta^* = \frac{a-\mu}{\sigma^2}$. We then get

$$I(a) = \frac{(a-\mu)^2}{2\sigma^2}$$

Which is indeed what we found using Gaussian tail bounds. □

- ***Proof of Cramer's Theorem***: We will carry out this proof in two steps
  - ***Step 1***: establish an upper bound

$$\limsup_n \frac{1}{n} \log \mathbb{P}\left(S_n > na\right) \le -I(a)$$

  - ***Step 2***: establish a lower bound

$$\liminf_n \frac{1}{n} \log \mathbb{P}\left(S_n > na\right) \ge -I(a)$$

Fix $a > \mu$ and $\theta > 0$.

  - ***Step 1***: Recall from Markov's Inequality that for any $\alpha > 0$, we have

$$\mathbb{P}\left(X_1 > \alpha\right) \le \frac{\mathbb{E}(X_1)}{\alpha}$$

For this proof, we will need a more general form of Markov's Inequality.

Lemma (Generalized Form of Markov's Inequality): Let $g : \mathbb{R}_+ \to \mathbb{R}_+$ be an increasing function such that $\mathbb{E}\left|g(X_1)\right| < \infty$. We then have that

$$\mathbb{P}(X_1 > \alpha) \le \frac{\mathbb{E}\left(g(X_1)\right)}{g(\alpha)}$$

Proof: We have that

$$\begin{aligned}\mathbb{E}\left(g(X_1)\right) &\ge \mathbb{E}\left[g(X_1); X_1 > \alpha\right] \\ &\ge g(\alpha)\mathbb{P}(X_1 > \alpha)\end{aligned}$$

As required. ■

Now, consider that in our case. Since θ is strictly positive, we can write

$$\begin{aligned}\mathbb{P}\left(S_n > na\right) &= \mathbb{P}\left(\theta S_n > \theta na\right) \\ &= \mathbb{P}\left(\exp(\theta S_n) > \exp(\theta na)\right) \\ &\le \mathbb{E}\left(e^{\theta S_n}\right)e^{-\theta na}\end{aligned}$$

(This is often known as a Chernoff bound. Note that we have used the generalized form of Markov's Inequality in the last line). We then have

$$\begin{aligned}\mathbb{P}\left(S_n > na\right) &\le e^{-\theta na}\mathbb{E}\left(e^{\theta S_n}\right) \\ &= e^{-\theta na}\mathbb{E}\left(\exp\left[\theta \sum_{i=1}^n X_1\right]\right) \\ &= e^{-\theta na}\mathbb{E}\left(\prod_{i=1}^n \exp(\theta X_1)\right) \\ &= e^{-\theta na}\left[M(\theta)\right]^n \\ &= e^{-\theta na}e^{n\varphi(\theta)} \\ &= \exp\left(-n[\theta a - \varphi(\theta)]\right)\end{aligned}$$

Let us try to optimize the bound (ie: make it as tight as possible). To do this, we choose θ such that

$$\theta \in \arg\sup_{\theta > 0}\left\{\theta a - \varphi(\theta)\right\}$$

Note, however, that

- The derivative of our objective, at $\theta = 0$, is given by $a - \varphi'(0) = a - \mu > 0$.
- We will show, below that our objective is concave.

Together, these two points imply that extending our optimization problem to $\theta \in \mathbb{R}$ will not change our optimum, since it necessarily lies in the positive quadrant. Thus, we can choose θ such that

$$\theta \in \arg\sup_{\theta \in \mathbb{R}} \left\{ \theta a - \varphi(\theta) \right\}$$

This is precisely the optimization problem involved in finding $I(a)$. Going back to the above, we therefore find that

$$\mathbb{P}\left(S_n > na\right) \le \exp\left(-nI(a)\right)$$

Precisely as requested. All that remains to do is to prove that $\varphi(\theta)$ is a convex function. Fix $\alpha \in (0,1)$ and $\theta_1, \theta_2 \in \mathbb{R}$. We then have

$$\begin{aligned} \varphi\left(\alpha\theta_1 + [1-\alpha]\theta_2\right) &= \log\left(M[\alpha\theta_1 + (1-\alpha)\theta_2]\right) \\ &= \log\left(\mathbb{E}[e^{\alpha\theta_1 X_1} e^{(1-\alpha)\theta_2 X_1}]\right) \\ &= \log\left(\mathbb{E}[e^{\alpha\theta_1 X} e^{(1-\alpha)\theta_2 X}]\right) \end{aligned}$$

?????????

- ***Step 2:*** We now proof the lower bound, in a very different way. Assume, for notational convenience, that X has a density $p(x)$ – this is not, however, required. Then, fix $\theta > 0$ and define a new density

$$p_\theta(x) = \frac{e^{\theta x} p(x)}{M(\theta)} \ge 0$$

which is a bona-fide density. Suppose $g : \mathbb{R} \to \mathbb{R}$ is such that $\mathbb{E}\left|g(X_1)\right| < \infty$, then

$$\mathbb{E}\left[g(X)\right] = \int g(x) p(x) \,\mathrm{d}x = \int g(x) \frac{p(x)}{p_\theta(x)} p_\theta(x) \,\mathrm{d}x = \mathbb{E}_\theta\left[g(X_1)\frac{p(x)}{p_\theta(x)}\right]$$

Similarly,

$$\mathbb{E}\left[g(X_1, \cdots, X_n)\right] = \mathbb{E}_\theta\left[g(X_1,\cdots,X_n)\frac{p(X_1,\cdots,X_n)}{p_\theta(X_1,\cdots,X_n)}\right]$$

Where here, $g:\mathbb{R}^n \to \mathbb{R}$. This is, effectively, a change of measure, and p / p_θ is a likelihood ratio.

Now, consider

$$\begin{aligned}
\mathbb{P}\left(S_n > na\right) &= \mathbb{E}\left[\mathbb{I}_{\{S_n > na\}}\right] \\
&= \mathbb{E}_\theta\left[\mathbb{I}_{\{S_n > na\}} \frac{p(X_1,\cdots,X_n)}{p_0(X_1,\cdots,X_n)}\right] \\
&= \mathbb{E}_\theta\left[\mathbb{I}_{\{S_n > na\}} \prod_{i=1}^n \frac{p(X_i)}{p_\theta(X_i)}\right] \\
&= \mathbb{E}_\theta\left[\mathbb{I}_{\{S_n > na\}} \prod_{i=1}^n \frac{p(X_i)}{e^{\theta X_i} p(X_i) / M(\theta)}\right] \\
&= \mathbb{E}_\theta\left[\mathbb{I}_{\{S_n > na\}} \frac{\left[M(\theta)\right]^n}{\exp\left(\theta \sum_{i=1}^n X_i\right)}\right] \\
&= \mathbb{E}_\theta\left[\mathbb{I}_{\{S_n > na\}} \exp\left(-\theta S_n + n\varphi(\theta)\right)\right]
\end{aligned}$$

Now, fix a $\delta > 0$

$$\begin{aligned}
\mathbb{P}\left(S_n > na\right) &\geq \mathbb{E}_\theta\left[\mathbb{I}_{\left\{na < S_n \leq na + \delta\sqrt{n}\right\}} \exp\left(-\theta S_n + n\varphi(\theta)\right)\right] \\
&\geq \mathbb{E}_\theta\left[\mathbb{I}_{\left\{na < S_n \leq na + \delta\sqrt{n}\right\}} \exp\left(-\theta\left(na + \delta\sqrt{n}\right) + n\varphi(\theta)\right)\right] \\
&= e^{-n(\theta a - \varphi(\theta))} e^{-\theta\delta\sqrt{n}} \mathbb{P}_\theta\left(na < S_n \leq na + \delta\sqrt{n}\right)
\end{aligned}$$

This is starting to look like what we want – it is crucial, however, to ensure that the last probability stays bounded away from 0 – if not, it tends to negative infinity. Unfortunately, according to our *original* distribution, it *does* fall to 0. Our hope is to choose our new distribution (characterized by θ) to ensure we get what we want.

Note that

$$\begin{aligned}
\mathbb{E}_\theta(X) &= \int \frac{x e^{\theta x} p(x)}{M(\theta)} \, \mathrm{d}x \\
&= \frac{M'(\theta)}{M(\theta)} \\
&= \varphi'(\theta)
\end{aligned}$$

Now, suppose there exists a θ^* such that $\varphi'(\theta)$ [in fact, we have assumed X_1 has a moment generating function that exists over the real line, so there *does* exist such a θ][2]. Then the mean of S_n under our new distribution would indeed be na, which gives us hope that the probability above will be bounded away from 0. Formally, by the central limit theorem, under θ^*

$$\frac{S_n - na}{\sqrt{n}} \Rightarrow \sigma^2 N(0,1)$$

if $(\sigma^*)^2 < \infty$, where $(\sigma^*)^2 = \mathbb{E}_{\theta^*}(X^2) - a^2$. Under the conditions of Cramer's Theorem, it turns out this is also satisfied.

We then have that

$$\begin{aligned}\mathbb{P}_{\theta^*}\left(na < S_n \le na + \delta\sqrt{n}\right) &= \mathbb{P}_{\theta^*}\left(0 < \frac{S_n - na}{\sigma^*\sqrt{n}} \le \frac{\delta}{\sigma^*}\right) \\ &\to \mathbb{P}\left(0 \le Z \le \frac{\delta}{\sigma^*}\right) > 0\end{aligned}$$

As required. Now, returning to our expression above,

$$\frac{1}{n}\log \mathbb{P}\left(S_n > na\right) = -\left[\theta^* a - \varphi(\theta^*)\right] - \frac{\theta^*\delta}{\sqrt{n}} + \frac{1}{n}\log \mathbb{P}_{\theta^*}\left(na < S_n \le na + \delta\sqrt{n}\right)$$

Taking a lim inf,

$$\liminf_n \frac{1}{n}\log \mathbb{P}\left(S_n > na\right) \ge -\left[\theta^* a - \varphi(\theta^*)\right]$$

The last thing we now need to verify is that the θ^* required is indeed the maximizer of the function above, to get $I(a)$. To do that, note that $\sup_{\theta\in\mathbb{R}}\left\{\theta a - \varphi(\theta)\right\}$ is a concave program, which implies first-order conditions are necessary and sufficient. In this case, they are $a = \varphi'(\theta)$, precisely as desired. ■

- ***Applications of Large Deviations Theory***

---

[2] We mentioned that it is enough to the moment generating function to exist in the neighborhood of 0, "and some extra conditions". These extra conditions are precisely those that ensure such a θ does exist.

- ***Example***: Let $X_1, X_2, \cdots$ be IID $N(\mu, \sigma^2)$ variables variables. Fix $a > \mu$. We saw, earlier that

$$\begin{aligned}\mathbb{P}\left(S_n > na\right) &\leq \exp\left(-nI(a)\right) \\ &= \exp\left(-n\tfrac{(a-\mu)^2}{2\sigma^2}\right)\end{aligned}$$

The central limit theorem allows us to "handwave" $S_n \approx n\mu + \sigma\sqrt{n}N(0,1)$. Using this result, however, we can be more exact – we can find a sequence a_n such that S_n eventually lies below $n\mu + a_n$ almost surely. Consider:

$$\begin{aligned}\mathbb{P}\left(S_n > n\mu + a_n\right) &= \mathbb{P}\left(S_n > n(\mu + \tfrac{a_n}{n})\right) \\ &\leq \exp\left(-n\frac{(a_n / n)^2}{2\sigma^2}\right)\end{aligned}$$

We now need to choose an a_n that ensure our envelope is summable. Let us consider, for $\delta > 0$, $a_n = \sqrt{2(1+\delta)\sigma^2 n \log n}$. This is very close to $\sqrt{n}$, which we might expect to work given the central limit theorem. With that choice of envelope,

$$\begin{aligned}\mathbb{P}\left(S_n > n\mu + a_n\right) &\leq \exp\left(-(1+\delta)\log n\right) \\ &= n^{-(1+\delta)}\end{aligned}$$

This is indeed summable. Thus, by the first Borrel-Cantelli Lemma, $\left\{S_n \leq n\mu + a_n\right\}$ almost surely. Using a similar kind of argument, we could show that $\left\{S_n \geq n\mu - a_n\right\}$ eventually almost surely. This is a much more precise statement than that of the central limit theorem. □

- ***Moderate deviations theory***. We saw that a "typical" deviation is of order $\sqrt{n}$, a "large" deviation is of order n. What about deviations in between?

  Theorem (Moderate Deviations): Under the conditions of Cramer's Theorem, for any sequence α_n such that

  - $\frac{\alpha_n}{\sqrt{n}} \to \infty$
  - $\frac{\alpha_n}{n} \to 0$

  for any $a > 0$,

$$\frac{n}{\alpha_n^2}\log \mathbb{P}\left(S_n > n\mu + a\alpha_n\right) \to -\frac{a^2}{2\sigma^2}$$

Proof of upper bound: WLOG, set $\mu = 0$ and $a = 1$.

$$\varphi(0) = 0 \qquad \varphi'(0) = 0 \qquad \varphi''(0) = \sigma^2$$

We then obtain, as $\theta \to 0$,

$$\varphi(\theta) \sim \frac{\theta^2\sigma^2}{2}$$

(Recall that $f(x) \sim g(x) \Leftrightarrow \lim \frac{f(x)}{g(x)} = 1$). Let us now use a Chernoff bound to deduce that

$$\begin{aligned}\mathbb{P}\left(\tfrac{S_n}{n} > \tfrac{\alpha_n}{n}\right) &\leq \exp(\theta \tfrac{\alpha_n}{n})\mathbb{E}\left[\exp(\theta \tfrac{S_n}{n})\right] \\ &= e^{n\varphi(\frac{\theta}{n})}e^{-\theta\frac{\alpha_n}{n}}\end{aligned}$$

But $\varphi(\frac{\theta}{n}) \sim \frac{\theta^2\sigma^2}{2n^2}$, so

$$\mathbb{P}\left(\tfrac{S_n}{n} > \tfrac{\alpha_n}{n}\right) \leq C\exp\left(\frac{\theta^2\sigma^2}{2n} - \theta\frac{\alpha_n}{n}\right)$$

Optimizing, we find

$$\mathbb{P}\left(\tfrac{S_n}{n} > \tfrac{\alpha_n}{n}\right) \leq C\exp\left(-\frac{\alpha_n^2}{2n\sigma^2}\right)$$

As required. ■

Lecture 5 – *16th February 2011*

Random Walks & Martingales

- ***Random walks***
  - ***Definition*** (Random Walk): A random process $\{S_n : n \geq 0\}$ is called a *random walk* (RW) if it can be represented as $S_n = \sum_{i=1}^n X_i$, with the $\{X_n\}$ independently and identically distributed, and independent of S_0.

    Remarks: If $S_0 = 0$:
    - $\mathbb{E}(S_n) = n\mathbb{E}(X_1)$ and $\mathbb{V}\text{ar}(S_n) = n\mathbb{V}\text{ar}(X_1)$
    - By the Strong Law of Large Numbers, $S_n \approx n\mathbb{E}(X_1)$
    - $S_n \approx n\mathbb{E}(X_1) + \sigma\sqrt{n}N(0,1)$

- ***Question***: Suppose T is a positive, integer valued random variable. Is it the case that $\mathbb{E}(S_T) = \mathbb{E}(T)\mathbb{E}(X_1)$?
- Let us consider the situation in which T is independent of the $\{S_n\}$ (which is equivalent to saying T is independent of the $\{X_n\}$). We then have

$$\mathbb{E}(S_T) = \mathbb{E}\left(\sum_{i=1}^T X_i\right) = \mathbb{E}\left(\mathbb{E}\left[\sum_{i=1}^T X_i \mid T\right]\right) = \mathbb{E}\left(T\mathbb{E}[X_1]\right) = \mathbb{E}(T)\mathbb{E}(X_1)$$

  Unfortunately, this answer is not particularly interesting or useful, simply because in all "interesting" situations, T is allowed to depend on the behavior of the system. Let us consider some more interesting cases.
- ***Example***: Consider a random walk defined as follows

$$X_i \overset{\text{IID}}{\sim} \begin{cases} 1 & \text{with prob } p \\ -1 & \text{with prob } 1-p \end{cases} \quad p > \tfrac{1}{2} \qquad\qquad S_n = \sum_{i=1}^n X_i$$

  In this case, $\frac{S_n}{n} \to 2p - 1 > 0$ a.s., and so $S_n \to \infty$ almost surely. Now, imagine $S_0 = -1$ and define

$$T = \sup\left\{n \geq 0 : S_n = 0\right\}$$

  Because $S_n \to \infty$, we have $\mathbb{P}(T < \infty) = 1$. Furthermore, by definition, $S_T = 0 \Rightarrow \mathbb{E}(S_T) = 0$. However, we have that $\mathbb{E}(X_1) = 2p - 1 > 0$ and that $\mathbb{E}(T) \geq 1$. Thus, it is clear, in this case that

$$\mathbb{E}(S_T) \neq \mathbb{E}(T)\mathbb{E}(X_1)$$

  Clearly, therefore, there can be some issues. In this case, the problem is that the time T is *anticipatory*; it is determined by full knowledge of the future of the random walk.
- ***Definition*** (Filtration): We previously defined a probability space by the triplet $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathcal{F}$ represented possible *events*. We define $\mathcal{F}_n = \sigma\left(X_0, \cdots, X_n\right)$ to be the *sigma algebra* generated by $X_0, \cdots, X_n$. This sigma algebra constrains all the *information* that is available from knowing the value of the variables $X_0, \cdots, X_n$. We then say that $\{\mathcal{F}_n\}$ is a *filtration*, and clearly, $\cdots \subseteq \mathcal{F}_n \subseteq \mathcal{F}_{n+1} \subseteq \cdots$.

- ***Definition*** (Stopping time): Suppose T is a non-negative integer-valued random variable. Then T is said to be a *stopping time* with respect to an underlying sequence $\{X_n\}$ if, for each $k \geq 0$,

$$\mathbb{I}_{\{T=k\}} = f_k\left(X_0, \cdots, X_k\right)$$

Where f_k is a deterministic function. In other words, we require

$$\{T = k\} \in \mathcal{F}_k \qquad \forall k$$

- ***Example*** (Hitting Times): Let $T = \inf\{n \geq 0 : X_n \in A\}$. We then have

$$\mathbb{I}_{\{T=k\}} = \mathbb{I}_{\{X_0 \notin A, \cdots, X_{k-1} \notin A, X_k \in A\}}$$

Similarly, we would define $T = \inf\{n \geq 0 : S_n \in A\}$, and we would then have

$$\mathbb{I}_{\{T=k\}} = \mathbb{I}_{\{S_0 \notin A, \cdots, S_{k-1} \notin A, S_k \in A\}}$$

- ***Proposition*** (Wald's First Identity): Let S_n be a random walk $S_n = \sum_{i=1}^n X_i$, with $S_0 = 0$, and let T be a stopping time with respect to the sequence $\{\mathcal{F}_n\}$ (where $\mathcal{F}_n = \sigma(X_1, \cdots, X_n)$)
  - If $X_i \geq 0$, then $\mathbb{E}(S_T) = \mathbb{E}(T)\mathbb{E}(X_1)$ [this could, of course, we infinity].
  - If $\mathbb{E}\left|X_i\right| < \infty$ and $\mathbb{E}(T) < \infty$, then $\mathbb{E}(S_T) = \mathbb{E}(T)\mathbb{E}(X_0)$

Proof: Let $S_T = \sum_{i=1}^T X_i = \sum_{i=1}^\infty X_i \mathbb{I}_{\{i \leq T\}}$. We then have

$$\mathbb{E}(S_T) = \mathbb{E}\left(\sum_{i=1}^\infty X_i \mathbb{I}_{\{i \leq T\}}\right)$$

Now, let us do both parts:

- ***First part**:* $X_i \geq 0$, and indicators are always positive, so by Fubini I, we can interchange the expectation and the sum:

$$\mathbb{E}(S_T) = \sum_{i=1}^\infty \mathbb{E}\left(X_i \mathbb{I}_{\{i \leq T\}}\right)$$

Consider, however, that

$$\begin{aligned}\mathbb{I}_{\{T \geq i\}} &= 1 - \mathbb{I}_{\{T < i\}} \\ &= 1 - \mathbb{I}_{\{T \leq i-1\}}\end{aligned}$$

This implies that $\mathbb{I}_{\{T \geq i\}} \in \mathcal{F}_{i-1}$. Going back to our sum

$$\begin{aligned}\mathbb{E}(S_T) &= \sum_{i=1}^{\infty} \mathbb{E}\left[\mathbb{E}\left(X_i \mathbb{I}_{\{i \le T\}} \mid \mathcal{F}_{i-1}\right)\right] \\ &= \sum_{i=1}^{\infty} \mathbb{E}\left[\mathbb{I}_{\{i \le T\}} \mathbb{E}\left(X_i \mid \mathcal{F}_{i-1}\right)\right] \\ &= \sum_{i=1}^{\infty} \mathbb{E}\left[\mathbb{I}_{\{i \le T\}} \mathbb{E}\left(X_1\right)\right] \\ &= \mathbb{E}(X_1) \sum_{i=1}^{\infty} \mathbb{E}\left(\mathbb{I}_{\{i \le T\}}\right)\end{aligned}$$

Using Fubini I again:

$$\begin{aligned}\mathbb{E}(S_T) &= \mathbb{E}(X_1)\mathbb{E}\left(\sum_{i=1}^{\infty} \mathbb{I}_{\{i \le T\}}\right) \\ &= \mathbb{E}(X_1)\mathbb{E}(T)\end{aligned}$$

- ***Second part:*** In this case, consider that, by the triangle inequality

$$\left|S_T\right| \le \sum_{i=1}^{T} \left|X_i\right|$$

By part 1, however,

$$\mathbb{E}\left(\sum_{i=1}^{T} \left|X_i\right|\right) = \mathbb{E}(T)\mathbb{E}\left|X_1\right| < \infty$$

The result then follows by Fubini II. ■

- ***Example*** (Gambler's Ruin): Consider a random walk in which $S_0 = 0$ and $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}$. We let $Z_n = k + S_n$, with $0 < k < N < \infty$ – intuitively, k is our initial wealth, and Z_n is our wealth at time n. Let

$$\begin{aligned}T &= \inf\left\{n \ge 0 : Z_n = 0 \text{ or } Z_n = N\right\} \\ &= \inf\left\{n \ge 0 : S_n = -k \text{ or } S_n = N - k\right\}\end{aligned}$$

Consider that $\mathbb{E}(Z_T) = k + \mathbb{E}(S_T)$. But consider also that, by definition

$$\begin{aligned}\mathbb{E}(Z_T) &= 0\mathbb{P}(Z_T = 0) + N\mathbb{P}(Z_T = N) \\ &= N\mathbb{P}(Z_T = N)\end{aligned}$$

Now, unfortunately, the X_i are not non-negative. However, $\left|X_i\right| \le 1$, and so it clearly has finite expectation. Let us now *assume* $\mathbb{E}(T) < \infty$ (this will be proved in homework 2). By Wald I, we then have

$$\mathbb{E}(S_T) = \mathbb{E}(X_1)\mathbb{T}(T) = 0$$

And so

$$\begin{aligned}N\mathbb{P}(Z_T = N) &= k \\ \mathbb{P}(Z_T = N) &= k / N\end{aligned}$$

- ***Proposition*** (Wald II): Suppose $\left\{X_i\right\}$ are IID *bounded* random variables, and $\mathbb{E}(X_1)=0$, $\mathbb{E}(X_1^2)=\sigma^2$. Let T be a stopping time with respect to $\left\{\mathcal{F}_n\right\}$, and such that $\mathbb{E}(T^2)<\infty$. Then
$$\mathbb{V}\mathrm{ar}(S_T)=\sigma^2\mathbb{E}(T)$$
Remark: If T were independent of the X_i, the result would be obvious. We seek to extend this to stopping times.

Proof: Clearly, under the assumptions of this proportion, Wald I holds – as such, $\mathbb{E}(S_T)=\mathbb{E}(T)\mathbb{E}(X)=0$. Now
$$\begin{aligned}\mathbb{E}(S_T^2)&=\mathbb{E}\left(\sum_{i=1}^T X_i\right)^2\\&=\mathbb{E}\left(\sum_{i=1}^T X_i^2\right)+2\mathbb{E}\left(\sum_{i=1}^{T-1}\sum_{j=i+1}^T X_iX_j\right)\end{aligned}$$
But by Wald I,
$$\begin{aligned}\mathbb{E}\left(\sum_{i=1}^T X_i^2\right)&=\mathbb{E}(T)\mathbb{E}(X_1^2)\\&=\sigma^2\mathbb{E}(T)\end{aligned}$$
Now, consider the second term
$$\mathbb{E}\left(\sum_{i=1}^{T-1}\sum_{j=i+1}^T X_iX_j\right)=\mathbb{E}\left(\sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty} X_iX_j\mathbb{I}_{\{T-1\geq i\}}\mathbb{I}_{\{j\leq T\}}\right)$$
Now, imagine it were possible to apply Fubini 2. We then have
$$\mathbb{E}\left(\sum_{i=1}^{T-1}\sum_{j=i+1}^T X_iX_j\right)=\sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty}\mathbb{E}\left(X_iX_j\mathbb{I}_{\{T-1\geq i\}}\mathbb{I}_{\{j\leq T\}}\right)$$
But note that because of the way we have written these sums, $T\geq j\Rightarrow T>i$. As such, we have that $\mathbb{I}_{\{T\geq j\}}\mathbb{I}_{\{T\geq i+1\}}=\mathbb{I}_{\{T\geq j\}}$. As such
$$\begin{aligned}\mathbb{E}\left(\sum_{i=1}^{T-1}\sum_{j=i+1}^T X_iX_j\right)&=\sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty}\mathbb{E}\left(X_iX_j\mathbb{I}_{\{T\geq j\}}\right)\\&=\sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty}\mathbb{E}\left[\mathbb{E}\left(X_iX_j\mathbb{I}_{\{T\geq j\}}\mid\mathcal{F}_{j-1}\right)\right]\\&=\sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty}\mathbb{E}\left[X_i\mathbb{I}_{\{T\geq j\}}\mathbb{E}\left(X_j\mid\mathcal{F}_{j-1}\right)\right]\\&=\sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty}\mathbb{E}\left[X_i\mathbb{I}_{\{T\geq j\}}\cdot 0\right]\\&=0\end{aligned}$$
Now, let us justify the use of Fubini 2. X is bounded, and so $\left|X\right|\leq K$

$$\begin{aligned} S_T^2 &= \left(\sum_{i=1}^T X_i\right)^2 \\ &\le \left(\sum_{i=1}^T |X_i|\right)^2 \\ &\le (KT)^2 = K^2T^2 \end{aligned}$$

This has a finite expectation, since T has a second moment. ■

- **Example** (Gambler's Ruin): In this case, $\mathbb{E}(X_1^2) = 1 = \sigma^2$, and we assume $\mathbb{E}(T^2) < \infty$ (again, see homework 2), we then have

$$\mathbb{E}(Z_T^2) = 0\mathbb{P}\left(Z_T = 0\right) + N^2\mathbb{P}\left(Z_T = N\right) = kN$$

However, we also have

$$\begin{aligned} \mathbb{E}(Z_T^2) &= \mathbb{E}(k + S_T)^2 \\ &= k^2 + 2k\mathbb{E}(S_T) + \mathbb{E}(S_T^2) \\ &= k^2 + \sigma^2\mathbb{E}(T) \\ &= k^2 + \mathbb{E}(T) \end{aligned}$$

And so

$$\mathbb{E}(T) = kN - k^2 = k(N-k)$$

- ***Martingales***
  - **Definition** (Martingale): A process $\left\{M_n : n \ge 0\right\}$ is called a *martingale* (MG) with respect to a filtration $\left\{\mathcal{F}_n\right\}$ if
    - $M_n \in \mathcal{F}_n$
    - $\mathbb{E}\left|M_n\right| < \infty \qquad \forall n$
    - $\mathbb{E}\left(M_n \mid \mathcal{F}_{n-1}\right) = M_{n-1}$
  - **Example** (Random walk): Let $M_n = S_n$, with $M_0 = 0$. Then
    - $S_n \in \mathcal{F}_n$
    - Provided $\mathbb{E}\left|X_i\right| < \infty$, then $\mathbb{E}\left|S_n\right| \le n\mathbb{E}\left|X_1\right| < \infty$
    - $\mathbb{E}\left[S_n \mid \mathcal{F}_{n-1}\right] = \mathbb{E}\left[S_{n-1} + X_n \mid \mathcal{F}_{n-1}\right] = S_{n-1} + \mathbb{E}(X_n)$

    Thus, if we let $\mathbb{E}(X_n) = 0$, then our random walk is a martingale.
  - **Example** (Variance martingale): Consider $M_n = S_n^2 - n\sigma^2$, with $\mathbb{E}(X_1^2) < \infty$ and $\mathbb{E}(X_1) = 0$. Then
    - Condition 1 holds.
    - $\mathbb{E}\left|M_n\right| \le \mathbb{E}(S_n^2) + n\sigma^2 = 2n\sigma^2 < \infty \quad \forall n$

- $\mathbb{E}\left(M_n \mid \mathcal{F}_{n-1}\right) = \mathbb{E}\left(S_n^2 - n\sigma^2 \mid \mathcal{F}_{n-1}\right) = \mathbb{E}\left(\left[S_{n-1} + X_n\right]^2 - n\sigma^2 \mid \mathcal{F}_{n-1}\right) = S_{n-1}$

- Consider that if we let $D_n = M_n - M_{n-1}$, we can write $M_n = \sum_{i=1}^n D_i + M_0$.
- ***Proposition***: Let $\{M_n : n \geq 0\}$ be a martingale with respect to $\{\mathcal{F}_n\}$ and put $D_i = M_i - M_{i-1}$. Then
  - $\mathbb{E}(D_i) = 0 \qquad \forall i$
  - If $\sup_{i\geq 1} \mathbb{E}(D_i^2) < \infty$, then $\mathbb{E}(D_i D_j) = 0 \quad \forall i \neq j$

  Remarks

  - Note that this theorem *almost* achieves the representation of a general martingale as a random walk. Indeed, it expresses our martingale as a sum of mean-0 uncorrelated random variables. The D_i, however, need not be independent (and indeed, in most "typical" cases, they will not be).
  - As we will see later, the boundedness condition is required to be able to argue that why exactly is this needed?

  $$\mathbb{E}\left|D_i D_j\right| \leq \sqrt{\mathbb{E}\left|D_i\right|^2}\sqrt{\mathbb{E}\left|D_j\right|^2} < \infty$$

  (Note: in this case, it would be enough to require the second moment of each difference to be bounded, but it seems less clunky to require the slightly more stringent uniform boundedness condition).

  Proof: Fix $i \geq 1$:

  $$\begin{aligned}\mathbb{E}(D_i) &= \mathbb{E}\left(M_i - M_{i-1}\right) \\ &= \mathbb{E}\left[\mathbb{E}(M_i - M_{i-1} \mid \mathcal{F}_{n-1})\right] \\ &= \mathbb{E}\left[\mathbb{E}(M_i \mid \mathcal{F}_{n-1}) - M_{i-1}\right] \\ &= 0\end{aligned}$$

  Now, fix $j > i \geq 1$:

  $$\begin{aligned}\mathbb{E}(D_i D_j) &= \mathbb{E}\left[D_i(M_j - M_{j-1})\right] \\ &= \mathbb{E}\left[\mathbb{E}\left(D_i(M_j - M_{j-1}) \mid \mathcal{F}_{j-1}\right)\right] \\ &= \mathbb{E}\left[D_i \mathbb{E}\left(M_j - M_{j-1} \mid \mathcal{F}_{j-1}\right)\right] \\ &= 0\end{aligned}$$

  (Note that this can be used to prove $\mathbb{V}\text{ar}(M_i) = \mathbb{V}\text{ar}(M_0) + \sum_{i=1}^n \mathbb{V}\text{ar}(D_i)$). ■

- ***Example***: Let $\{X_n\}$ be IID, with $\mathbb{E}(X_i) = 1$ and $\mathbb{E}|X_i| < \infty$. And let $M_n = \prod_{i=1}^n X_i$. This is a martingale:
  - Condition 1 satisfied.
  - $\mathbb{E}|M_n| = \prod_{i=1}^n \mathbb{E}|X_i| < \infty$
  - Conditioning

$$\begin{aligned}\mathbb{E}\left(M_n \mid \mathcal{F}_{n-1}\right) &= \mathbb{E}\left(\prod_{i=1}^n X_i \mid \mathcal{F}_{n-1}\right) \\ &= \mathbb{E}\left(X_n \prod_{i=1}^{n-1} X_i \mid \mathcal{F}_{n-1}\right) \\ &= M_{n-1}\mathbb{E}\left(X_n \mid \mathcal{F}_{n-1}\right) \\ &= M_{n-1}\end{aligned}$$

It is quite astounding, therefore, that even this process can be written as the sum of uncorrelated increments!

- ***Optional Stopping Theorem for Martingales***
  - ***Question:*** If T is a stopping time, when is it true that $\mathbb{E}(M_T) = \mathbb{E}(M_0)$? This is the question that will be concerning us in this section. Let us consider some simple examples.
  - ***Example***: Let $M_n = S_n - n\mu$, with $M_0 = 0$ and $\mathbb{E}(X_1) = \mu$. When is it the case that $\mathbb{E}(M_T) = \mathbb{E}(M_0) = 0$? Effectively, we are asking when it is the case that

$$\mathbb{E}(S_T) = \mu \mathbb{E}(T)$$

This is precisely the subject matter of Wald's First Identity. ☐

  - ***Expample***: Let $M_n = S_n^2 - n\sigma^2$, with $\mathbb{E}(X_1) = 0, \mathbb{E}(X_1^2) = \sigma^2 < \infty$. When is it true that $\mathbb{E}(M_T) = \mathbb{E}(M_0) = 0$? Effectively, we are asking when it is the case that

$$\mathbb{E}(S_T^2) = \sigma^2 \mathbb{E}(T)$$

This is precisely the subject matter of Wald's Second Identity. ☐

  - ***Proposition***: Let $\{M_n : n \geq 0\}$ be a martingale with respect to $\{\mathcal{F}_n\}$, and let T be a stopping time. Then for each $m \geq 1$

$$\mathbb{E}\left(M_{T \wedge m}\right) = \mathbb{E}\left(M_0\right)$$

Where $T \wedge m = \min(T, m)$.

Lecture 6 – *24th February 2011*

- ***Proof***: Let $\{D_i\}$ be the martingale differences, and write

$$M_{T\wedge n} = \sum_{i=1}^{T\wedge n} D_i + M_0 = \sum_{i=1}^{n} D_i \mathbb{I}_{\{T\geq i\}} + M_0$$

Take expectations

$$\mathbb{E}(M_{T\wedge n}) = \sum_{i=1}^{n} \mathbb{E}\left(D_i \mathbb{I}_{\{T\geq i\}}\right) + \mathbb{E}\left(M_0\right)$$

We know, however, that $\mathbb{I}_{\{T\geq i\}} \in \mathcal{F}_{i-1}$

$$\begin{aligned}\mathbb{E}(M_{T\wedge n}) &= \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{E}\left(D_i \mathbb{I}_{\{T\geq i\}} \mid \mathcal{F}_{i-1}\right)\right] + \mathbb{E}\left(M_0\right) \\ &= \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{I}_{\{T\geq i\}}\mathbb{E}\left(D_i \mid \mathcal{F}_{i-1}\right)\right] + \mathbb{E}\left(M_0\right) \\ &= \mathbb{E}\left(M_0\right)\end{aligned}$$

As required. ■

Remark: Consider that

- Consider that $\lim_{n\to\infty} M_{T\wedge n} \to M_T$. As such, $\mathbb{E}\left[\lim_{n\to\infty} M_{T\wedge n}\right] = \mathbb{E}\left[M_T\right]$.
- By the theorem above, however, $\mathbb{E}\left[M_{T\wedge n}\right] = \mathbb{E}\left[M_0\right]$, which implies than $\lim_{n\to\infty} \mathbb{E}\left[M_{T\wedge n}\right] = \mathbb{E}\left[M_0\right]$.

Thus, every optional stopping theorem boils down to the following interchange argument – if we can make the interchange, then $\mathbb{E}\left[M_T\right] = \mathbb{E}\left[M_0\right]$:

$$\boxed{\mathbb{E}\left[\lim_{n\to\infty} M_{T\wedge n}\right] = \lim_{n\to\infty} \mathbb{E}\left[M_{T\wedge n}\right]}$$

- ***Corollary I***: Let $\left(M_n : n \geq 0\right)$ be a martingale with respect to $\{\mathcal{F}_n\}$ and T be a stopping time such that T is bounded (in other words, there exists a $K < \infty$ such that $\mathbb{P}\left(T < K\right) = 1$), then $\mathbb{E}(M_T) = \mathbb{E}(M_0)$.
- ***Corollary II***: Let $\left|M_{T\wedge n}\right| \leq Z$, with $\mathbb{E}(Z) < \infty$. Then by dominated convergence, the interchange holds and $\mathbb{E}(M_T) = \mathbb{E}(M_0)$.
- ***Corollary III***: Let $\left(M_n : n \geq 0\right)$ be a martingale with respect to $\{\mathcal{F}_n\}$ and T be a stopping time such that $\mathbb{E}(T) < \infty$. Provided the martingale differences are uniformly bounded ($\mathbb{E}\left[\left|D_i\right| \mid \mathcal{F}_{i-1}\right] \leq C < \infty$), then $\mathbb{E}(M_T) = \mathbb{E}(M_0)$.

Remark: This is a stronger version of Wald I. There, we had essentially established the OST under the conditions $\mathbb{E}\left|X_1\right| < \infty$, $\mathbb{E}(T) < \infty$. In the case of a random walk, the X are exactly the martingale differences. This theorem is slightly stronger because here, it is enough for the *conditional* increments to be bounded).

Proof: We have

$$M_{T\wedge n} = \sum_{i=1}^{T\wedge n} D_i + M_0$$

M_0 is integrable. Now consider that,

$$\begin{aligned}\left|\sum_{i=1}^{T\wedge n} D_i\right| &\leq \sum_{i=1}^{T\wedge n}\left|D_i\right| \\ &= \sum_{i=1}^{\infty}\left|D_i\right|\mathbb{I}_{\{T\geq i\}}\end{aligned}$$

Let us now take expectations. By Fubini I, we can then swap the expectation and sum, since the summands are positive

$$\mathbb{E}\sum_{i=1}^{\infty}\left|D_i\right|\mathbb{I}_{\{T\geq i\}} = \sum_{i=1}^{\infty}\mathbb{E}\left[\left|D_i\right|\mathbb{I}_{\{T\geq i\}}\right] = \sum_{i=1}^{\infty}\mathbb{E}\left[\mathbb{I}_{\{T\geq i\}}\mathbb{E}\left[\left|D_i\right| \mid \mathcal{F}_{n-1}\right]\right]$$

Since the martingale differences are uniformly bounded bt C, this is $\leq C\mathbb{E}(T)$. As such

$$\left|M_{T\wedge n}\right| \leq C\mathbb{E}(T) + \left|M_0\right|$$

M_0 is integrable, and by the statement of the theorem, $\mathbb{E}(T) < \infty$. As such, our martingale is bounded by an integrable variable, and the OST holds by corollary II. ■

- ### *Martingale Limit Theory & Concentration Inequalities*
  - In Optional Stopping Theory, we were looking for results about *random* times T. We now consider *very large* times T.
  - ***Proposition (Martingale Convergence Theorem)***: Let $\left\{M_n : n \geq 0\right\}$ be a martingale with respect to $\left\{\mathcal{F}_n\right\}$ such that $\sup_{n\geq 1}\mathbb{E}\left|M_n\right| < \infty$ (note: this is a much stronger condition than $\mathbb{E}\left|M_n\right| < \infty$ for all n, because each item could be finite but the sum might diverge). Then

$$M_n \to M_\infty \qquad \text{a.s.}$$

for some finite bounded random variable M_∞.

Proof: See Williams, page 109

- Recall that $M_n = M_0 + \sum_{i=1}^n D_i$. This is a random-walk like structure, even though the D_i may not be independent. We can write

$$\frac{M_n}{n} = \frac{M_0}{n} + \frac{1}{n}\sum_{i=1}^n D_i$$

At an intuitive level, since the D_i are uncorrelated, we might expect the SLLN to hold and the last sum to converge to 0. If we succeed in showing that, we have effectively shown that M_n/n converges to 0.

- ***Proposition (Martingale SLLN)***: Let $\{M_n : n \geq 0\}$ be a martingale with respect to $\{\mathcal{F}_n\}$ such that $\sup_{n\geq 1} \mathbb{E}(D_n^2) < \infty$ (this, as we saw above, is a requirement for the increments to be uncorrelated). Then

$$\frac{M_n}{n} \to 0 \qquad \text{a.s.}$$

Proof: Put $\tilde{M}_n = \sum_{k=1}^n \frac{D_k}{k}$ (where, as usual $D_k = M_k - M_{k-1}$. Clearly, $\tilde{M}_n \in \mathcal{F}_n$, and

$$\begin{aligned}
\mathbb{E}\left[\tilde{M}_n \mid \mathcal{F}_{n-1}\right] &= \mathbb{E}\left[\sum_{k=1}^{n-1} \tfrac{D_k}{k} + \tfrac{D_n}{n} \mid \mathcal{F}_{n-1}\right] \\
&= \sum_{k=1}^{n-1} \tfrac{D_k}{k} + \frac{1}{n}\mathbb{E}\left[D_n \mid \mathcal{F}_{n-1}\right] \\
&= \tilde{M}_{n-1}
\end{aligned}$$

and $\mathbb{E}\left|\tilde{M}_n\right| < \infty$ for all n, due to the conditions on the second moments of D_n.. Thus, it is a martingale. Now

$$\mathbb{E}\left|\tilde{M}_n\right|^2 = \mathbb{E}\left[\sum_{k=1}^n \frac{D_k^2}{k^2}\right] + \mathbb{E}\left[\sum_{i\neq j}^n \frac{D_i D_j}{ij}\right]$$

The second term is equal to 0, since the D are uncorrelated.

$$\mathbb{E}\left|\tilde{M}_n\right|^2 \leq \sum_{k=1}^n \frac{\sup_{n\geq 1}\mathbb{E}(D_k^2)}{k^2} < \infty$$

This implies that

$$\sup_{n\geq 1} \mathbb{E}\left|\tilde{M}_n\right|^2 < \infty \Rightarrow \sup_{n\geq 1} \mathbb{E}\left|\tilde{M}_n\right| < \infty$$

(Notice that it is often easier to work with second moments rather than first moments, despite the fact finiteness of second moments is a stronger condition than finiteness for first moments). If so, by the Martingale Convergence Theorem,

$$\tilde{M}_n \to \tilde{M}_\infty \qquad \text{a.s.}$$

Now, let

$$A = \left\{\omega : \sum_{k=1}^n \tfrac{D_k(\omega)}{k} \to M_\infty(\omega)\right\} \qquad \mathbb{P}(A) = 1$$

$$B = \left\{\omega : \frac{1}{n}\sum_{k=1}^n D_k(\omega) \to 0\right\}$$

By Kroenecker's Lemma $A \subseteq B$. Thus, $\mathbb{P}(B) = 1$. This proves our Theorem. ■

Remark: When we proved the SLLN, we went to great pains to work with a first moment only. In this case, we work with second moments, which makes the proof much simpler. It is interesting to note, however, that the scaling of D_k / k is "overkill". Our series would also have converged for $\frac{D_k}{k^{(1/2)+\delta}}$. Carrying the entire proof through, we obtain $M_n / n^{(1/2)+\delta} \to 0$ a.s. Thus, assuming more than first moments has indeed resulted in a stronger conclusion. Mapping this back to the IID world and letting $S_n = \sum_{i=1}^n X_i$ with $\mathbb{E}(X_1^2) < \infty$, we have obtained the stronger result that $\frac{1}{\sqrt{n(\log n)^{1+\delta}}} S_n \to 0$.

- ***Proposition (Central Limit Theorem for Martingales)***: Let $\{M_n : n \geq 0\}$ be a martingale with respect to $\{\mathcal{F}_n\}$ and put $V_n = \max_{1\leq i\leq n}\left|D_i\right|$. If
  - $\sup_n \frac{\mathbb{E}(V_n^2)}{n} < \infty$
  - $\left|V_n\right| / \sqrt{n} \Rightarrow 0$
  - $\frac{1}{n}\sum_{i=1}^n D_i^2 \Rightarrow \sigma^2$ (deterministic and finite). This makes our martingale "very similar" to a random walk.

  Then

$$\frac{1}{\sqrt{n}} M_n \Rightarrow \sigma N(0,1)$$

Remark: The third condition is key here. For a random walk $S_n = \sum_{i=1}^n Y_i$, it is the case that $\mathbb{V}\text{ar}(S_n) = n\mathbb{V}\text{ar}(Y_1)$. Similarly, for a martingale,

$\mathbb{V}\text{ar}(M_n) = \mathbb{V}\text{ar}(M_0) + \sum_{i=1}^n \mathbb{V}\text{ar}(D_k)$ – we want to ensure these two results are as close to each other as possible.

- ***Proposition (Azuma-Hoeffding Inequality)***: Let $\{M_n : n \geq 0\}$ be a martingale with respect to $\{\mathcal{F}_n\}$ such that

$$\left|M_n - M_{n-1}\right| = \left|D_n\right| \leq c_n \qquad \forall n$$

Then, for all $n > 0, \lambda > 0$

$$\mathbb{P}\left(\left|M_n - M_0\right| \geq \lambda\right) \leq 2\exp\left(-\frac{\lambda^2}{2\sum_{k=1}^n c_k^2}\right)$$

Remark: Suppose that $\left|D_n\right| \leq c \ \forall k$, then we can re-write this as

$$\mathbb{P}\left(\left|M_n - M_0\right| \geq \lambda\right) \leq 2\exp\left\{-\frac{\lambda^2}{2c^2 n}\right\}$$
$$\mathbb{P}\left(\left|M_n - M_0\right| \geq xc\sqrt{n}\right) \leq 2\exp\left\{-\frac{x^2}{2}\right\}$$

Or, choosing $\lambda = \sqrt{\kappa n \log n}$, we obtain

$$\mathbb{P}\left(\left|M_n - M_0\right| \geq \sqrt{\kappa n \log n}\right) \leq 2\exp\left(-\frac{\kappa \log n}{2c}\right)$$

Choosing, for example, $\kappa = 4c$ produces a summable sequence, which can be used to obtain an almost sure result.

Proof[3]: Define $D_k = M_k - M_{k-1}$. Let $\theta > 0$, and write

$$D_k = -\tfrac{1}{2}c_k\left(1 - \tfrac{D_k}{c_k}\right) + \tfrac{1}{2}c_k\left(1 + \tfrac{D_k}{c_k}\right)$$

Both terms in parentheses are non-negative (since the martingale differences are bounded by c_k) and add up to one. As such, we can use Jensen's Inequality to write

$$\begin{aligned} e^{\theta D_k} &\leq \tfrac{1}{2}\left(1 - \tfrac{D_k}{c_k}\right)e^{-\theta c_k} + \tfrac{1}{2}\left(1 + \tfrac{D_k}{c_k}\right)e^{\theta c_k} \\ &= \frac{e^{-\theta c_k} + e^{\theta c_k}}{2} + \frac{D_k}{2c_k}\left(e^{\theta c_k} - e^{-\theta c_k}\right) \end{aligned}$$

[3] The proof of Azuma's Inequality was actually covered in the next lecture. For expositional purposes, we chose to present this material here instead.

Recall, however, that since M is a martingale, $\mathbb{E}\left[D_k \mid \mathcal{F}_{k-1}\right] = 0$. Recall, in addition, that $\frac{e^{-x}+e^{x}}{2} \le e^{x^2/2}$, which can be proved using a Taylor expansion. As such,

$$\begin{aligned}\mathbb{E}\left[e^{\theta D_k} \mid \mathcal{F}_{k-1}\right] &\le \frac{e^{-\theta c_k} + e^{\theta c_k}}{2} \\ &\le e^{(\theta c_k)^2/2}\end{aligned}$$

Now, consider

$$\begin{aligned}\mathbb{E}\left[e^{\theta(M_n - M_0)}\right] &= \mathbb{E}\left[\prod_{k=1}^n \exp(\theta D_k)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\prod_{k=1}^n \exp(\theta D_k) \mid \mathcal{F}_{n-1}\right]\right] \\ &= \mathbb{E}\left[\prod_{k=1}^{n-1} \exp(\theta D_k)\mathbb{E}\left[e^{\theta D_n} \mid \mathcal{F}_{n-1}\right]\right] \\ &\le e^{(\theta c_n)^2/2}\mathbb{E}\left[\prod_{k=1}^{n-1} \exp(\theta D_k)\right]\end{aligned}$$

Applying this process repeatedly, we get

$$\mathbb{E}\left[e^{\theta(M_n - M_0)}\right] \exp\left(\sum_{k=1}^n \frac{(\theta c_k)^2}{2}\right)$$

However,

$$\begin{aligned}\mathbb{P}\left(M_n - M_0 \ge \lambda\right) &= \mathbb{P}\left(\exp^{\theta(M_n - M_0)} \ge e^{\theta\lambda}\right) \\ &\le \mathbb{E}\left[e^{\theta(M_n - M_0)}\right] e^{-\theta\lambda} \\ &\le \exp\left(\frac{\theta^2}{2}\sum_{k=1}^n c_k^2 - \theta\lambda\right)\end{aligned}$$

Optimizing over θ to make the bound as tight as possible, we find $\theta^* = \lambda / \sum_{i=1}^n c_i^2$. Substituting back into the equation, we obtain the required result. ■

- ***Example***: Suppose the X_i are IID uniform random variables taking values in the set $\mathcal{N} = \left\{1, \cdots, N\right\}$. Let $B = (b_1, \cdots, b_k)$, with $b_i \in \mathcal{N}$ and $k \ll \left|\mathcal{N}\right| = n$. Now, draw n IID random copies of X, and write

$$X_1^n = \left(X_1, \cdots, X_n\right)$$

Let $k \ll n$. We wonder how many times the specific sequence B will appear in X_1^n. We will call this number $R_{n,k}$. Finding the distribution of this is difficult. However,

$$\mathbb{E}(R_{n,k}) = \overbrace{(n-k+1)}^{\substack{\text{Number of ways we} \\ \text{can "place" } B \text{ on } X_1^n \\ \text{by "sliding" it up and} \\ \text{down the } X_1^n}} \overbrace{\left(\tfrac{1}{N}\right)^k}^{\substack{\text{Probability of} \\ \text{finding } B \text{ at} \\ \text{a particular position}}}$$

Let

$$M_0 = \mathbb{E}(R_{n,k})$$

And let

$$M_i = \mathbb{E}\left[R_{n,k} \mid \mathcal{F}_i\right] \qquad \mathcal{F}_i = \sigma(X_1, \cdots, X_i)$$

This is then a martingale, and $M_n = R_{n,k}$. But consider that since each X can only be part of at most k of the inequalities

$$\left|M_i - M_{i-1}\right| = \left|\mathbb{E}[R_{n,k} \mid \mathcal{F}_i] - \mathbb{E}[R_{n,k} \mid \mathcal{F}_{i-1}]\right| \\ \leq k$$

Using the A-H inequality, we then have

$$\mathbb{P}\left(\left|M_n - M_0\right| > \lambda\right) \leq 2\exp\left(-\frac{\lambda^2}{2nk^2}\right)$$
$$\Leftrightarrow$$
$$\mathbb{P}\left(\left|R_{n,k} - \mathbb{E}(R_{n,k})\right| > \lambda\right) \leq 2\exp\left(-\frac{\lambda^2}{2nk^2}\right)$$
$$\Leftrightarrow$$
$$\mathbb{P}\left(\left|R_{n,k} - \mathbb{E}(R_{n,k})\right| > xk\sqrt{n}\right) \leq 2\exp\left(-\frac{x^2}{2}\right)$$

We can then construct a confidence interval

$$R_{n,k} \in \left[\mathbb{E}(R_{n,k}) - \chi_\alpha k\sqrt{n}, \mathbb{E}R_{n,k} + \chi_\alpha k\sqrt{n}\right]$$

Where χ_α is chosen to make the probability of the interval $= 1-\alpha$. □

- ***Example***: Let $\left\{X_n : n \geq 0\right\}$ be an irreducible, recurrent Markov Chain (MC) with transition matrix P. Then, the only bounded solution to

$$P\boldsymbol{f} = \boldsymbol{f}$$

is a constant vector (a multiple of $\boldsymbol{1}$). (Note: the state space could be infinite).

Remark: We can think of the vector $\boldsymbol{f}$ as a function $\boldsymbol{f} : \mathcal{S} \to \mathbb{R}$, where $\mathcal{S}$ is the set of states of the chain. Note as well that this is *not* the steady state equation, which satisfies $\boldsymbol{\pi} = \boldsymbol{\pi}^\top P$.

Proof: Let $M_n = \boldsymbol{f}(X_n)$ for an $\boldsymbol{f}$ that satisfies the assumption of the proposition. The main question here is one of uniqueness, since it is clear that a constant vectors can solve $P\boldsymbol{f} = \boldsymbol{f}$. Let us first show that M_n is a martingale:

- $\mathbb{E}\left|M_n\right| = \mathbb{E}\left|\boldsymbol{f}(X_n)\right| \leq c$
- $M_n \in \mathcal{F}_n = \sigma(X_1, \cdots, X_n)$ (in fact, it only depends on the last X_n).
- $\mathbb{E}\left[M_n \mid \mathcal{F}_{n-1}\right] = \mathbb{E}\left[\boldsymbol{f}(X_n) \mid X_{n-1}\right] = P\boldsymbol{f}(X_{n-1}) = \boldsymbol{f}(X_{n-1})$

As such, $\left\{M_n : n \geq 0\right\}$ is a martingale with respect to $\left\{\mathcal{F}_n\right\}$. Note that $\sup_{n\geq 1} \mathbb{E}\left|M_n\right| < \infty$, because we have assumed that $\boldsymbol{f}$ is bounded. Thus, $M_n \to M_\infty$ almost surely. Suppose there exists $x, y \in \mathcal{S}$ such that

$$\boldsymbol{f}(x) < \boldsymbol{f}(y)$$

Since $\left\{X_n : n \geq 0\right\}$ is irreducible and recurrent

$$\begin{aligned} &\liminf_n f(X_n) \leq \boldsymbol{f}(x) &&\text{a.s.} \\ &\limsup_n f(X_n) \geq \boldsymbol{f}(y) &&\text{a.s.} \end{aligned}$$

This means, however, that $\liminf_n \boldsymbol{f}(X_n) \neq \limsup_n \boldsymbol{f}(X_n)$, which contradicts the convergence statement. □

Lecture 7 – *2nd March 2011*

- ***Stochastic stability***[4]
  - ***Deterministic motivation***: Consider a dynamical system $X(t)$ for which $\mathrm{d}X(t) = f(X(t))\,\mathrm{d}t$, with $f : \mathbb{R} \to \mathbb{R}$ (this can be thought of as the "equation of motion" of the system).

    Now, consider an "energy" function $g : \mathbb{R} \to \mathbb{R}_+$, with

$$\frac{\mathrm{d}g\left(X(t)\right)}{\mathrm{d}t} \leq -\varepsilon \qquad \varepsilon > 0$$

[4] Some parts of this topic were covered at the end of the previous lecture. For expositional purposes, we chose to present the material here instead.

This is effectively a statement of the fact the energy of the system is "forced" down to 0, since it is "constantly decreasing" As such, $\exists t_0(x,\varepsilon)$ such that $g\left(X(t)\right) = 0$ for all $t \geq t_0$; in other words, our dynamical process is "pushed" towards a "stable" state.

- We now need to adapt this idea to a discrete stochastic process. We can write our "equation of motion" as $X_{n+1} - X_n = f(X_n)$. To add stochasticity, we can write $X_{n+1} = \psi(X_n, \varepsilon_{n+1})$, where the ε_i are random variables. This is effectively the definition of a Markov process, provided ε_n is independent of X_0. If we make the ε_i IID, the Markov process becomes time-homogeneous.

  Now consider the "energy" function – it would be too strong to ask for the energy of the process to decrease along *every* path. We therefore require it to decrease in expectation:

$$\mathbb{E}\left[g(X_{n+1}) - g(X_n) \mid X_n\right] \leq -\varepsilon$$

  Since we need this be true *whatever state* our process first starts in, we can write $\mathbb{E}_x[\,\cdot\,] = \mathbb{E}[\,\cdot \mid X_0 = x]$, and the condition above becomes

$$\mathbb{E}_x\left[g(X_1)\right] - g(x) \leq -\varepsilon$$

- We now specialize this to a particular stochastic process. Consider a Markov chain $\left\{X_n : n \geq 0\right\}$ that is irreducible. We would like to know whether the chain has a steady state. For a finite state space, all we need is to check for solutions to $\boldsymbol{\pi}^\top P = \boldsymbol{\pi}, \boldsymbol{\pi}^\top \boldsymbol{1} = 1$; indeed, the existence of such a $\boldsymbol{\pi}$ is associated with positive recurrence. If the spate space is countably infinity, however, things get slightly more complicated; the two equations above do not suffice, and we also require $\boldsymbol{\pi} \geq \boldsymbol{0}$, which makes things more complicated. Here, we attempt to find simpler conditions for stability to hold.

- ***Proposition***: Let $\left\{X_n : n \geq 0\right\}$ be an irreducible markov chain on a countable state space $\mathcal{S}^k$. Let $K \subseteq \mathcal{S}$ be a set containing a finite number of states. Then, if there exists a function $g : \mathcal{S} \to \mathbb{R}_+$ such that (recall $\mathbb{E}_x[\cdot] = \mathbb{E}[\cdot \mid X_0 = x]$)

$$\mathbb{E}_x g(X_1) - g(x) \leq -\varepsilon \qquad \forall x \in \bar{\mathcal{K}} \text{ and some } \varepsilon > 0$$

$$\mathbb{E}_x g(X_1) < \infty \qquad \forall x \in \mathcal{K}$$

Then $(X_n, n \geq 0)$ is positive recurrent

- ***Proof***: Let g be as in the proposition, and fix $x \in \mathcal{S}$. Construct

$$M_n = g(X_n) - g(x) - \sum_{k=1}^{n-1} (Ag)(X_k)$$

Where $(Ag)(X_k) = \mathbb{E}_{X_k} g(X_{k+1}) - g(X_k)$. Write

$$\begin{aligned} M_n &= g(X_n) - g(x) - \sum_{k=1}^{n-1} \mathbb{E}_{X_k}[g(X_{k-1})] - g(X_k) \\ &= -\sum_{k=1}^{n} \left[\mathbb{E}_{X_{k-1}} g(X_k) - g(X_k)\right] \\ &= -\sum_{k=1}^{n} D_k \end{aligned}$$

Now, defining $\mathcal{F}_n = \sigma\left(X_0, \cdots, X_n\right)$, we have that

$$\begin{aligned} \mathbb{E}\left[D_k \mid \mathcal{F}_{k-1}\right] &= \mathbb{E}\left\{\mathbb{E}_{X_{k-1}} g(X_k) - g(X_k) \mid \mathcal{F}_{k-1}\right\} \\ &= \mathbb{E}_{X_{k-1}} g(X_k) - \mathbb{E}\left\{g(X_k) \mid \mathcal{F}_{k-1}\right\} \\ &= \mathbb{E}_{X_{k-1}} g(X_k) - \mathbb{E}_{X_{k-1}} g(X_k) \\ &= 0 \end{aligned}$$

And so D_k is a martingale difference. Now, consider the set $K \subseteq \mathcal{S}$ in the proposition. Let $T_k = \inf\left\{n \geq 0 : X_n \in K\right\}$, and Fix $m \in \mathbb{N}$. $T_k \wedge m$ is a bounded stopping time so we can apply the OST.

$$\mathbb{E}\left(M_{T_k \wedge m}\right) = \mathbb{E}\left(M_0\right) = 0$$

By plugging this into the definition of the martingale M_n

$$\mathbb{E}\left[g(X_{T_k \wedge m})\right] - g(x) - \mathbb{E}\left[\sum_{k=0}^{(T_k \wedge m)-1} (Ag)(x_k)\right] = 0$$

If $x \in K^c$,

$$\mathbb{E}_x g(X_1) - g(x) = (Ag)(x) \leq -\varepsilon$$

However, consider the sum – up to the hitting time, all the summands will be outside K, and will therefore be less than or equal to $-\varepsilon$. As such

$$\sum_{k=0}^{(T_k \wedge m)-1} (Ag)(x_k) \leq -\varepsilon\left(T_k \wedge m\right)$$

And so

$$-\mathbb{E}\left[\sum_{k=0}^{(T_k \wedge m)-1} (Ag)(x_k)\right] \geq \varepsilon \mathbb{E}\left(T_k \wedge m\right)$$

Combining our two inequalities

$$\overbrace{\mathbb{E}\left[g(X_{T_k \wedge m})\right]}^{\leq 0} - g(x) = \mathbb{E}\left[\sum_{k=0}^{(T_k \wedge m)-1} (Ag)(x_k)\right]$$
$$\leq -\varepsilon \mathbb{E}_x\left[T_k \wedge m\right]$$
$$-g(x) \leq -\varepsilon \mathbb{E}_x\left[T_k \wedge m\right]$$

And so

$$\mathbb{E}_x\left[T_k \wedge m\right] \leq g(x) / \varepsilon$$

But $0 \leq T_k \wedge m \nearrow T_k$ and $T_k < \infty$. So by monotone convergence

$$\mathbb{E}_x T_K \leq g(x) / \varepsilon \qquad\qquad x \in K^c$$

There is therefore a finite upper bound over the expectation – the mean return time to the set K is therefore finite in expectation. ■

- "***Application***": Consider the stochastic system $X_{n+1} = \alpha X_n + Z_{n+1}$ with $\left\{Z_n\right\}$ iid and $X_0 = x$. We want conditions on α and distributions of Z such that X_n is stable. Use

$$g(x) = \left|x\right|$$
$$\mathbb{E}_x g(X_1) = \mathbb{E}_x\left|\alpha x + Z_1\right| \leq \left|\alpha\right|\left|x\right| + \mathbb{E}\left|Z_1\right|$$

So if $\mathbb{E}\left|Z_1\right| < \infty$ and $\left|\alpha\right| < 1$, then we might be able to make this work, because we would then have

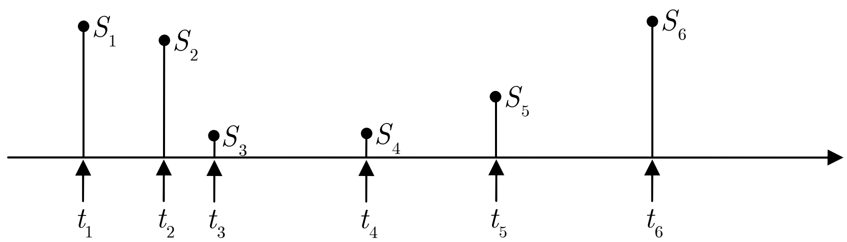
$$\mathbb{E}_x g(X_1) - g(x) \leq \left(\left|\alpha\right| - 1\right)\left|x\right| + \mathbb{E}(Z) \leq -\varepsilon \qquad \text{provided } \left|X\right| \geq \frac{\mathbb{E}\left|Z\right| + \varepsilon}{1 - \left|\alpha\right|}$$

Queuing

- ***Sample path methods***
  - We will consider a simple queuing model:
    - One single buffer, with unlimited space.
    - FIFO (first-in-first-out) service
    - A single server which processes work at unit rate.

  This model is often called the $G/G/1$ queuing model; arrival times are arbitrarily distributed, workloads are arbitrarily distributed (neither necessarily independent) and there is a single server.

- The variability in the system comes from the flow of work into the system. The points in our sample space are $\omega = \left\{(t_n, S_n) : n \in \mathbb{Z}\right\}$, with $S_n \geq 0$ and both S_n and t_n finite. A single point in the sample space is an infinite set of these numbers. Here is a pictorial representation of a single point, $\omega \in \Omega$ in our sample space:



- We assume the flow is stationary – in other words, $\theta_\tau \omega =_{\text{dist}} \omega$, for any $\omega \in \Omega$, where $\theta_\tau \omega = (t + \tau, S)$ and $\omega = (t, S)$.
- ***Definition (load)***: We define the *load* of the system as $\rho = \mathbb{E}\left(\sum_{n\in\mathbb{Z}} S_n \mathbb{I}_{t_n \in [0,1]}\right)$. The positioning of the interval [0,1] does not matter, since the flow is stationary. If the incoming load is greater than 1, the system will eventually saturate; if it is less than 1, it'll rest sometimes. The case $\rho = 1$ needs separate analysis.
- We also assume the flows are ergodic, and that

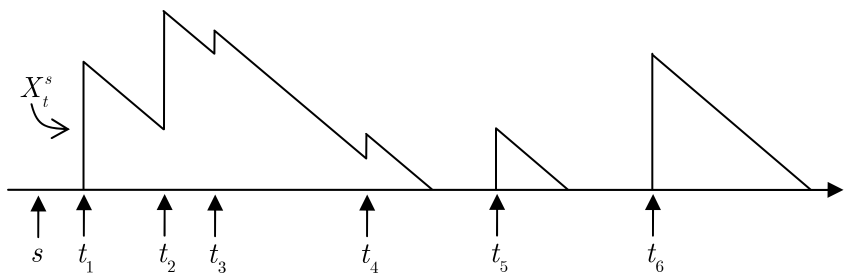
$$\frac{\sum_{n\in\mathbb{Z}} S_n \mathbb{I}_{t_n \in [s,t]}}{t - s} \begin{aligned} &\to_{s\to-\infty} \rho \\ &\to_{t\to\infty} \rho \end{aligned}$$

- Let $W_t(\omega)$ denote the work in the system at time t for a point ω in the sample space. We will simply denote this as W_t. The time between "empty times" at the server is called a *busy cycle*.
- Now

$$W_t(\omega) = \begin{cases} \alpha & t = 0 \\ (\alpha - t)^+ & 0 \leq t \leq t_1 \\ (\alpha - t)^+ + S_1 & t = t_2 \\ (W_{t_n} - t - t_n)^+ & t \in [t_n, t_{n+1}) \\ \left(W_{t_n} - (t_{n-2} - t_n)\right)^+ + S_{n+1} & t = t_{n+1} \end{cases}$$

These are called Lindley's Euqations.

- Set $s < t$ and $\omega \in \Omega$. Define $X_t^s(\omega)$ to be the work in the system at time t given the system is empty at time s. For the point ω depicted above (and assuming $s < t_1$), X_t^s would look like this



- For $s' < s$, it is clear that

$$X_t^{s'}(\omega) \geq X_t^s(\omega)$$

Because this variable is non-decreasing in s, $X_t^*(\omega) = \lim_{s \to -\infty} X_t^s(\omega)$ exists. This is the "steady state" of the system; what we would observe if we had started the system a very, very long time in the past.

- Now, define a time-shift operator θ_τ so that $X_t^s(\theta_\tau \omega) = X_{t+\tau}^{s+\tau}(\omega)$. As such, we have

$$\lim_{s \to -\infty} X_t^s(\theta_\tau \omega) = \lim_{s \to -\infty} X_{t+\tau}^{s+\tau}(\omega) = X_{t+\tau}^*(\omega) = X_t^*(\theta_\tau \omega)$$

As such, $X_t^*(\omega) =_d X_t^*(\theta_\tau \omega)$. Shift invariance also holds for the starred process.

Lecture 8 – *23rd March 2011*

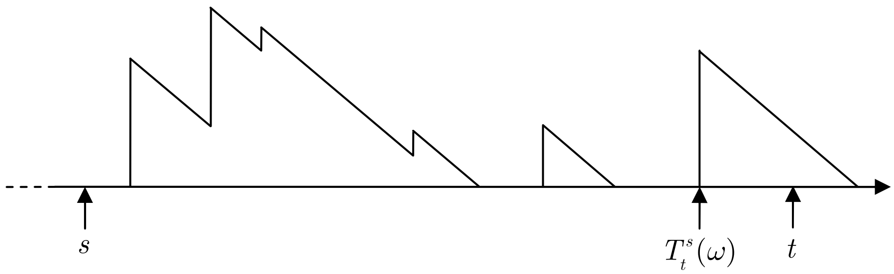
- We have seen a number of properties of X_t^*, including the fact that it exists. It might, however, be equal to infinity. We now look for conditions under which X_t^* is finite.

- ***Proposition***: If $\rho < 1$, then $X_t^* < \infty$ a.s. for all t.

Proof: Fix $t \in \mathbb{R}$ and $\omega \in \Omega$ and define

$$T_t^s(\omega) = \sup\left\{\tau < t : X_\tau^s(\omega) = 0\right\}$$

Intuitively, this is the last empty time before time t (this is clearly not a stopping time; just a random time):



Now, consider that (dropping the argument for X_t^s for simplicity)

$$\begin{aligned} X_t^s &= \overbrace{X_{T_t^s}^s}^{\substack{\text{Work in system}\\ \text{at } T_t^s}} + \overbrace{\sum_{n\in\mathbb{Z}} S_n \mathbb{1}_{\{t_n \in [T_t^s, t]\}}}^{\substack{\text{Work entering the}\\ \text{system in } [T_t^s, t]}} - \overbrace{(t - T_t^s)}^{\substack{\text{Work processed}\\ \text{in } [T_t^s, t]}} \\ &= \sum_{n\in\mathbb{Z}} S_n \mathbb{1}_{\{t_n \in [T_t^s, t]\}} - (t - T_t^s) \\ &\le \sum_{n\in\mathbb{Z}} S_n \mathbb{1}_{\{t_n \in [T_t^s, t]\}} \end{aligned}$$

Now clearly, for $s' < s$, $T_t^{s'} \le T_t^s$; in other words, T_t^s decreases as s decreases. This is because by moving s (empty time) further in the past, we are potentially introducing more work into the system). This implies that $\lim_{s\to-\infty} T_t^s = T_t^{-\infty}$ exists. Let us assume that $T_t^{-\infty} > -\infty$. This implies that there is some finite time in the past at which the system was empty, which implies that the amount of work at t, X_t^*, is finite.

All we therefore need to do, therefore is to show that $\rho < 1 \Rightarrow T_t^{-\infty} > -\infty$ a.s. We do this by contradiction; suppose $\rho < 1$ and $T_t^{-\infty} = -\infty$ on a set ω of sample paths. From the second line above, we have,

$$X_t^s = \sum_{n\in\mathbb{Z}} S_n \mathbb{1}_{\{t_n \in [T_t^s, t]\}} - (t - T_t^s)$$

Now, however, that $X_t^s \ge 0$ (there can never be negative work in the system), and so, dividing by $t - T_t^s$ throughout, we obtain

$$\frac{\sum_{n\in\mathbb{Z}} S_n \mathbb{1}_{\{t_n \in [T_t^s, t]\}}}{t - T_t^s} \ge 1$$

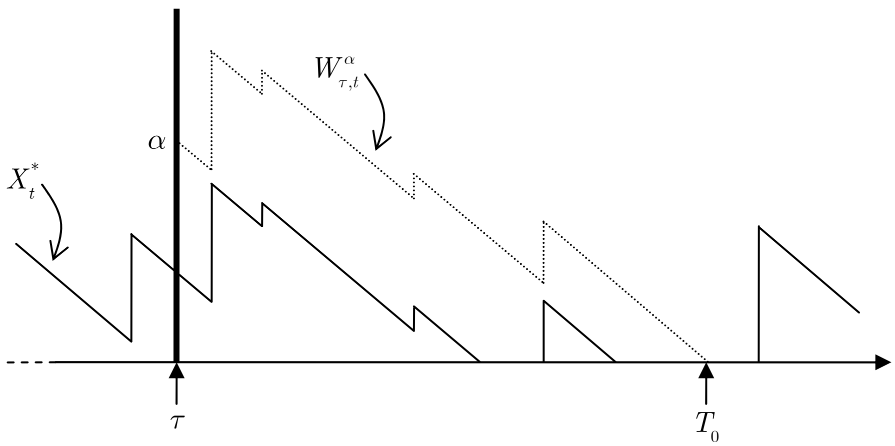
Note that if we were to replace T_t^s with s, the LHS of this expression would form a sequence with limit to ρ as $s \to -\infty$. By the definition of a limit, however, this is also true for any *subsequence* of that sequence. But since we have assumed

$T_t^s \to -\infty$ as $s \to -\infty$, the LHS above is such a subsequence. Thus, letting $s \to -\infty$, we find

$$1 \le \frac{\sum_{n\in\mathbb{Z}} S_n \mathbb{1}_{\{t_n \in [T_t^s, t]\}}}{t - T_t^s} \to_{s\to-\infty} \rho$$

This is a contradiction, since we have assumed $\rho < 1$. As such, $T_t^{-\infty} > -\infty$ a.s. and $X_t^* < \infty$ a.s. ■

- ***Definition (Coupling Time)***: Let $W_{\tau,t}^\alpha(\omega)$ denote the load in a system, at time t and along sample path ω, given that the system is allowed to start with load α at time τ. As before, let $X_t^*(\omega)$ denote the load in the system at time t given that system started empty at $t = -\infty$ (the "steady state"). Then T_0, called the *coupling time*, is the first time at which both the paths "couple" – past that point, the two paths behave identically.



- ***Proposition***: If $\rho < 1$, then $T_0 < \infty$ a.s.

  Proof: WLOG, assume $\tau = 0$ and denote $W_{\tau,t}^\alpha \equiv W_t^\alpha$. Also assume, WLOG, that $\alpha > X_\tau^*(\omega)$ (ie: the transient process starts *higher* than the steady-state process, as pictured above). In this case, the diagram above should make it clear that the coupling time is the first time at which both processes are empty. Since the transient process is more loaded than the steady-state process, this is effectively the time at which the transient process first empties. In other words,

$$T_0 = \inf\left\{t > 0 : W_t^\alpha = 0\right\}$$

Assume $\rho < 1$ and suppose $T_0 = \infty$ on a set of nonzero probability. This implies that $W_t^\alpha > 0 \ \forall t \geq 0$ and that the server is *constantly* working for all $t \geq 0$. As such

$$0 \leq W_t^\alpha = \overbrace{\alpha}^{\substack{\text{Work at} \\ t=0}} + \overbrace{\sum\nolimits_{n\in\mathbb{Z}} S_n \mathbb{1}_{\{t_n \in [0,t]\}}}^{\text{New work in } [0,t]} - \overbrace{t}^{\substack{\text{Work processed} \\ \text{in } [0,t]}}$$

Dividing by t and re-arranging, we obtain

$$1 \leq \frac{\alpha}{t} + \frac{\sum\nolimits_{n\in\mathbb{Z}} S_n \mathbb{1}_{\{t_n \in [0,t]\}}}{t} \to_{t\to\infty} \rho$$

This is a contradiction, and so $T_0 < \infty$ a.s. ■

- ***Example***: This theorem comes in particularly useful, for example, in estimating quantities like $\mathbb{P}\left(X_t^* \in A\right)$. Consider the following estimator, based on a queue started at an arbitrary point α

  $$\hat{p}_t(A) = \frac{1}{t}\int_0^t \mathbb{1}_{\{W_s^\alpha \in A\}} \ \mathrm{d}s$$

  As t grows, one might expect this to be a good estimator for $\mathbb{P}\left(X_t^* \in A\right)$. To show this formally, consider that

  $$\begin{aligned} \hat{p}_t(A) &= \frac{1}{t}\int_0^{T_0} \mathbb{1}_{\{W_s^\alpha \in A\}} \ \mathrm{d}s + \frac{1}{t}\int_{T_0}^t \mathbb{1}_{\{W_s^\alpha \in A\}} \ \mathrm{d}s \\ &\leq \frac{T_0}{t} + \frac{1}{t}\int_{T_0}^t \mathbb{1}_{\{X_s^* \in A\}} \ \mathrm{d}s \\ &\to 0 + \mathbb{P}\left(X_t^* \in A\right) \end{aligned}$$

  We can bound the estimator the other way by simply ignoring the first term. Either way, we find that $\hat{p}_t(A) \to \mathbb{P}\left(X_t^* \in A\right)$ a.s. as $t \to \infty$. Our estimator is therefore consistent. Intuitively, this is because the chains eventually couple. □
- We now ask *how fast* this convergence occurs…
- ***Proposition***: If $\rho < 1$

  $$\sup_{A_1,\cdots,A_n} \left| \mathbb{P}\left\{W_{t_1+\tau}^\alpha \in A_1, \cdots, W_{t_n+\tau}^\alpha \in A_n\right\} - \mathbb{P}\left\{X_{t_1}^* \in A_1, \cdots, X_{t_n}^* \in A_n\right\}\right| \to_{\tau\to\infty} 0$$

  Where $A_1, \cdots, A_n$ are measurable sets and for all $(t_1, \cdots, t_n) \in \mathbb{R}^n$ and $\alpha > 0$. This is called *convergence is total variation* and is stronger than weak convergence.

Proof: We prove the theorem when $n = 2$. This generalizes straightforwardly. Fix $\alpha > 0$, $t_1, t_2 \in \mathbb{R}$, $\tau > 0$ and sets A_1 and A_2. Now

$$\begin{aligned}&\left|\mathbb{P}\left\{W^\alpha_{t_1+\tau} \in A_1, W^\alpha_{t_2+\tau} \in A_2\right\} - \mathbb{P}\left\{X^*_{t_1} \in A_1, X^*_{t_2} \in A_2\right\}\right| \\ &\quad = \left|\mathbb{P}\left\{W^\alpha_{t_1+\tau} \in A_1, W^\alpha_{t_2+\tau} \in A_2\right\} - \mathbb{P}\left\{X^*_{t_1+\tau} \in A_1, X^*_{t_2+\tau} \in A_2\right\}\right|\end{aligned}$$

Calling the first event $\mathcal{A}_\tau$ and the second $\mathcal{B}_\tau$, we can write this as

$$\left|\mathbb{P}\left(\mathcal{A}_\tau, T_0 > \tau\right) + \mathbb{P}\left(\mathcal{A}_\tau, T_0 \leq \tau\right) - \mathbb{P}\left(\mathcal{B}_\tau, T_0 > \tau\right) - \mathbb{P}\left(\mathcal{B}_\tau, T_0 \leq \tau\right)\right|$$

Note, however, that if $T_0 \leq \tau$, the steady-state process is identical to the transient process. As such, the second and fourth terms cancel, and we can write this as

$$\begin{aligned}\left|\mathbb{P}\left(\mathcal{A}_\tau, T_0 > \tau\right) - \mathbb{P}\left(\mathcal{B}_\tau, T_0 > \tau\right)\right| &\leq \max\left\{\mathbb{P}\left(\mathcal{A}_\tau, T_0 > \tau\right), \mathbb{P}\left(\mathcal{B}_\tau, T_0 > \tau\right)\right\} \\ &\leq \mathbb{P}\left(T_0 > \tau\right) \\ &\to_{\tau\to\infty} 0\end{aligned}$$

Where the last step follows from the fact $T_0 < \infty$ a.s. if $\rho < 1$. Since this result does not depend on the sets A_1, A_2, we can take a supremum over all such sets and obtain our required result. ■

- We can compare to what happens in Markov chains. Consider a finite-state, irreducible Markov chain, and consider the quantity

$$\begin{aligned}\left|\mathbb{P}\left\{X_n = i \mid X_0 = j\right\} - \mathbb{P}_\pi\left\{X_n = i\right\}\right| &\leq \max_{i,j} \mathbb{P}_j\left\{T_i > n\right\} \\ &\leq \left(\max_{i,j} \mathbb{E}_j e^{\theta T_i}\right) e^{-\theta n}\end{aligned}$$

Here, the first probability is taken with respect to a Markov chain that starts in an arbitrary state j (the transient process). The second probability is taken with respect to the steady state. Since the expectation of the moments of T is finite (see homework 2), this implies that this difference falls exponentially fast.

- ***Proposition***: If $\rho > 1$, then for all $\alpha > 0$, $\liminf_{t\to\infty} \frac{W^\alpha_t}{t} > 0$ almost surely. In other words, the workload increases linearly with time.

Proof: Denote the cumulative idle time up to time t as I_t. We then have

$$\begin{aligned}W^\alpha_t &= \alpha + \sum_{n\in\mathbb{Z}} S_n \mathbb{1}_{\{t_n \in [0,t]\}} - \left(t - I_t\right) \\ &\geq \sum_{n\in\mathbb{Z}} S_n \mathbb{1}_{\{t_n\in[0,t]\}} - t\end{aligned}$$

Dividing by t

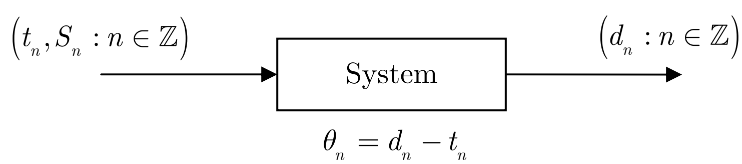
$$\frac{W_t^\alpha}{t} \ge \frac{\sum_{n\in\mathbb{Z}} S_n \mathbb{1}_{\{t_n \in [0,t]\}}}{t} - 1$$

Letting $t \to \infty$, we get that

$$\liminf_{t\to\infty} \frac{W_t^\alpha}{t} \ge \rho - 1 > 0$$

This is intuitively sensible – the amount that accumulates in the system is whatever load there is in excess of 1. ■

- ***Little's Law (Conservation Laws)***
  - We now consider a setting in which work $\left(t_n, S_n : n \in \mathbb{Z}\right)$ enters a system and then leaves the system at a time $\left(d_n : n \in \mathbb{Z}\right)$. We let θ_n be the *sojourn time* of the n^{th} job in the system, given by $d_n - t_n$.



We do not put any constraints on the working of the system. For example, the server could spend some of its time idling. In particularly, we do *not* assume that the server processes work on a FIFO basis, so it is not necessarily the case that $d_1 < d_2 < \cdots$.

We define the following quantities

$$\begin{aligned} A(t) &= \sup\left\{n : t_n \le t\right\} = \text{number of arrivals in } [0,t] \\ D(t) &= \text{number of of departures in } [0,t] \\ N(t) &= A(t) - D(t) = \text{work in the system at time } t \end{aligned}$$

The only two assumptions we make is that the following two statements are true for every sample path

  - $\lim_{t\to\infty} \dfrac{A(t)}{t} = \lambda \in (0,\infty)$
  - $\lim_{n\to\infty} \dfrac{\sum_{k=1}^{n} \theta_k}{n} = \overline{\theta} \in (0,\infty)$

  - ***Proposition***: Under these assumptions,

$$\frac{1}{t}\int_0^t N(s)\ \mathrm{d}s \to \lambda\overline{\theta}$$

Intuitively, this states that the average amount of work in the system at any given time is equal to the average number of arrivals per unit time multiplied by the average sojourn time.

Proof:

$$\int_0^t N(s)\ \mathrm{d}s = \int_0^t A(s) - D(s)\ \mathrm{d}s$$

Note that we can write

$$\begin{aligned} A(s) &= \sum\nolimits_{n\in\mathbb{Z}} \mathbb{1}_{\{t_n \le s\}} \\ D(s) &= \sum\nolimits_{n\in\mathbb{Z}} \mathbb{1}_{\{d_n \le s\}} \end{aligned}$$

As such

$$N(s) = \sum\nolimits_{n\in\mathbb{Z}} \mathbb{1}_{\{t_n \le s\}} - \mathbb{1}_{\{d_n \le s\}} = \sum\nolimits_{n\in\mathbb{Z}} \mathbb{1}_{\{t_n \le s \le d_n\}}$$

(The last equality follows because any jobs that arrive after s won't be counted at all, and any events that arrive and leave before s will be counted by both indicators and therefore cancel out). By swapping the summation and integration (valid by Fubini), we obtain

$$\begin{aligned} \int_0^t N(s)\ \mathrm{d}s &= \sum\nolimits_{n\in\mathbb{Z}} \int_0^t \mathbb{1}_{\{t_n \le s \le d_n\}}\ \mathrm{d}s \\ &= \sum\nolimits_{n\in\mathbb{Z}} \left\{\begin{matrix}\text{Amount of time job } n \text{ was} \\ \text{in the system during } [0,t]\end{matrix}\right\} \end{aligned}$$

We can bound this above by considering the sojourn time of all arrivals up to and including time t (though some of them may overrun past t) and lower bound it by considering the sojourn time of all job that depart before time t (even though some jobs that leave after t do spend *some* time in the system before t). This gives

$$\sum\nolimits_{i=0}^{D(t)} \theta_{n_i} \le \int_0^t N(s)\ \mathrm{d}s \le \sum\nolimits_{n=0}^{A(t)} \theta_n$$

Where n_i is the index of the i^{th} job to leave the system (since we have no assumed FIFO processing discipline, we cannot assume that $\theta_i = i$).

Before we continue, we will need the following claim

Claim: Under the two assumptions above

$$\frac{\theta_n}{t_n} \to 0 \qquad \text{as } n \to \infty$$

Intuitively, the second assumption states θ_n grows slower than n, so we would expect this to be true.

Proof: Consider that

$$\frac{\theta_n}{n} = \frac{\sum_{k=1}^{n} \theta_k - \sum_{k=1}^{n-1} \theta_k}{n} = \frac{1}{n}\sum_{k=1}^{n} \theta_k - \frac{1}{n}\sum_{k=1}^{n-1} \theta_k \to \bar{\theta} - \bar{\theta} \to 0$$

Consider also that

$$\frac{n}{t_n} = \frac{A(t_n)}{t_n} \to \lambda \text{ as } t_n \to \infty$$

Finally,

$$\frac{\theta_n}{n}\frac{n}{t_n} \to 0 \cdot \lambda = 0$$

As required. ■

Now by our claim, for all $\varepsilon > 0$, there exists an $N(\varepsilon)$ such that for $n > N(\varepsilon)$

$$\frac{\theta_n}{t_n} \le \varepsilon \Rightarrow \frac{d_n - t_n}{t_n} \le \varepsilon$$

This implies that $d_n \le (1+\varepsilon)t_n \ \forall n > N(\varepsilon)$. This means that all jobs after the $N(\varepsilon)^{\text{th}}$ job that arrive in $[0, t]$ will have departed by time $t(1+\varepsilon)$. Therefore

$$\sum_{n=N(\varepsilon)+1}^{A(t)} \theta_n \le \sum_{i=0}^{D\left(t(1+\varepsilon)\right)} \theta_{n_i}$$

Putting this together with the bounds developed above, we find that

$$\frac{1}{t(1+\varepsilon)}\sum_{n=N(\varepsilon)+1}^{A(t)} \theta_n \le \frac{1}{t(1+\varepsilon)}\int_0^{t(1+\varepsilon)} N(s)\ \mathrm{d}s \le \frac{1}{t(1+\varepsilon)}\sum_{n=1}^{A\left(t(1+\varepsilon)\right)} \theta_n$$

Let's first consider the upper bound

$$\frac{1}{t(1+\varepsilon)}\sum_{n=1}^{A\left(t(1+\varepsilon)\right)} \theta_n = \left(\frac{A\left(t(1+\varepsilon)\right)}{t(1+\varepsilon)}\right)\left(\frac{1}{A\left(t(1+\varepsilon)\right)}\sum_{n=1}^{A\left(t(1+\varepsilon)\right)} \theta_n\right) \to \lambda\bar{\theta}$$

The first term tends to λ, by assumption 1. The second term tends to $\bar{\theta}$ because

- By assumption 1, $A(t) \to_{t\to\infty} \infty$

- By assumption 2, $\frac{\sum \theta_n}{n} \to \overline{\theta}$, which means any subsequence thereof also $\to \overline{\theta}$. Since $A\left(t(1+\varepsilon)\right) \to \infty$, the second term above is precisely such a subsequence.

This implies that

$$\limsup_{t\to\infty} \frac{1}{t}\int_0^t N(s)\ \mathrm{d}s \le \lambda\overline{\theta}$$

Now the lower bound

$$\begin{aligned}\frac{1}{t(1+\varepsilon)}\sum_{n=N(\varepsilon)+1}^{A(t)} \theta_n &= \frac{1}{1+\varepsilon}\frac{A(t)}{t}\frac{1}{A(t)}\sum_{n=N(\varepsilon)+1}^{A(t)}\theta_n \\ &= \frac{1}{1+\varepsilon}\frac{A(t)}{t}\frac{\sum_{n=0}^{A(t)}\theta_n - \sum_{n=0}^{N(\varepsilon)}\theta_n}{A(t)} \\ &= \frac{1}{1+\varepsilon}\frac{A(t)}{t}\left(\frac{\sum_{n=0}^{A(t)}\theta_n}{A(t)} - \frac{\sum_{n=0}^{N(\varepsilon)}\theta_n}{A(t)}\right) \\ &\to \frac{\lambda\overline{\theta}}{1+\varepsilon}\end{aligned}$$

Again, the first term tends to λ by assumption 1. The second term tends to $\overline{\theta}$ by a similar logic as above, and by noting that since $N(\varepsilon) < \infty$, the second term in brackets tends to 0. As such, we get

$$\liminf_{t\to\infty} \frac{1}{t}\int_0^t N(\varepsilon)\ \mathrm{d}s \ge \frac{\lambda\overline{\theta}}{1+\varepsilon}$$

Since this is true for all $\varepsilon > 0$, this, together with the lim sup above, proves our theorem. ■

LECTURE 9 – *25th March 2011*

- ***Single-server queue; IID case***
  - We now specialize our analysis to a situation in which the workloads and inter-arrival times are IID. Letting $\tau_n = t_n - t_{n-1}$ be the time gap before the n^{th} job arrives, this situation requires the $\{S_n\}$ and $\{\tau_n\}$ to be IID. There is, once again, only one server. We often denote this situation $GI/GI/1$.
  - We also denote by w_n the time that the n^{th} job has to wait in queue before it is served. In that respect, $d_n = t_n + w_n + S_n$. Now, consider that

$$w_{n+1} = \begin{cases} 0 & d_n \le t_{n+1} \\ d_n - t_{n+1} & d_n > t_{n+1} \end{cases}$$
$$= \left(d_n - t_{n+1}\right)^+$$

As such

$$w_{n+1} = \left(w_n + S_n - \tau_{n+1}\right)^+$$

Let $Z_n = S_{n-1} - \tau_n$. We then have

$$w_{n+1} = \left(w_n + Z_{n+1}\right)^+ = \max\left(0, w_n + Z_{n+1}\right)$$

Since the Z_n are IID random variables, this is none other than a random walk "capped off" at the origin. Letting $\sigma_n = \sum_{j=1}^n Z_j$, we find that

$$\begin{aligned} w_1 &= \max\left(Z_1, 0\right) = \max\left(\sigma_1, 0\right) \\ w_2 &= \left(w_1 + Z_2\right)^+ \\ &= \sigma_2 + \max(0, -\sigma_1, -\sigma_2) \\ &= \sigma_2 - \min_{0 \le k \le 2} \sigma_k \end{aligned}$$

In fact, carrying this analysis forwards, we find that

$$w_n = \sigma_n - \min_{0 \le k \le n} \sigma_k$$

The second term takes into account the fact we have a *reflected* random walk, and prevents the walk from going negative.

- One thing the $GI/GI/1$ framework gives us over the $G/G/1$ framework is that we can now say something more about the distribution of the waiting times:

$$\begin{aligned} w_n &= \sigma_n - \min_{k \le n} \sigma_k \\ &= \max_{0 \le k \le n} \left\{\sigma_n - \sigma_k\right\} \\ &= \max_{0 \le k \le n} \left\{\sum_{j=k+1}^n Z_j\right\} \end{aligned}$$

Consider, however, that the Z_j are IID. As such, we can change indices on the Z at will and still maintain equality in distribution. Therefore

$$\begin{aligned} w_n &=_d \max_{0 \le k \le n} \left\{\sum_{j=1}^k Z_j\right\} \\ &= \max_{0 \le k \le n} \left\{\sigma_k\right\} \\ &= M_n \end{aligned}$$

Since M_n is a non-decreasing sequence, $M_n \to_{\text{a.s.}} M_\infty = \max_{k \ge 0} \sigma_k$. As such, $w_n \Rightarrow M_\infty$ as $n \to \infty$. [Note: convergence in distribution is the best we can hope for in this case, because M_n has some structure (the fact it's non-decreasing) that w_n does not. However, since this is a Markov chain, a stationary distribution is all we could really want].

- If the random walk has positive drift – in other words, if

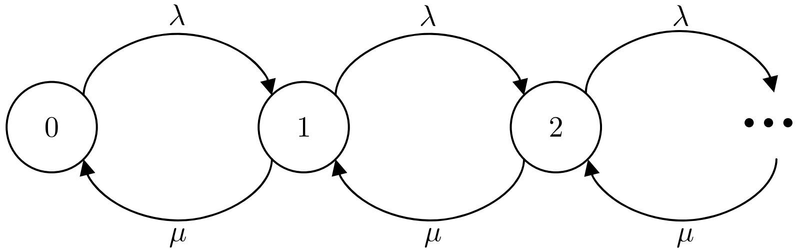
$$\mathbb{E}(Z_n) = \mathbb{E}(S_{n-1}) - \mathbb{E}(\tau_n) > 0$$

– then the random walk drifts to infinity and the waiting times get infinitely large. This is consistent with our findings in the $G/G/1$ queue, since $\mathbb{E}(\tau_{n+1}) > \mathbb{E}(S_n) \Leftrightarrow \rho < 1$. On the other hand, if the random walk has negative drift, the chain is stable and the waiting times return to 0 infinitely often almost surely (we motivated this result in homework 2 using a simpler reflected random walk).

- ***The single-server M/M/1 Markovian queue***
  - We now consider the most tractable of all single-server queue models; the $M/M/1$ in queue. In that case, we assume the $\{S_n\}$ are IID and exponentially distributed with parameter μ whereas the $\{\tau_n\}$ are IID and exponentially distributed with parameter λ.
  - Consider the process

$$X(t) = \text{Number of jobs in system at time } t \geq 0$$

$\{X(t) : t \geq 0\}$ is a continuous-time Markov Chain with countable state space. In fact, it is a birth-and-death process:



We now proceed to analyze this CTMC (note that $f(h) = o(h) \Rightarrow \lim_{h\to\infty} \frac{f(h)}{h} = 0$) Consider $t > 0$ and a small $h > 0$. We have

$$\mathbb{P}\left\{X(t+h)=j \mid X(t)=i\right\}=\begin{cases}\lambda h+o(h) & j=j+1 \\ \mu h+o(h) & j=j-1 \\ 1-(\lambda+\mu)h+o(h) & j=i \\ o(h) & \text{otherwise}\end{cases}$$

Our transition matrix P then takes the form

$$P_h=\begin{bmatrix} 1-\lambda h & \lambda h & 0 & & \\ \mu h & 1-(\lambda+\mu)h & \lambda h & 0 & \\ 0 & \mu h & 1-(\lambda+\mu)h & \lambda h & 0 \\ & 0 & \mu h & \ddots & \ddots \\ & & 0 & \ddots & \ddots \end{bmatrix}$$

We can write this as

$$P_h=I+\begin{bmatrix} -\lambda & \lambda & 0 & & \\ \mu & -(\lambda+\mu) & \lambda & 0 & \\ 0 & \mu & -(\lambda+\mu) & \lambda & 0 \\ & 0 & \mu & \ddots & \ddots \\ & & 0 & \ddots & \ddots \end{bmatrix}h+o(h)$$

$$P_h=I+Qh+o(h)$$

Q is the rate matrix, defined by $Q=\lim_{h\downarrow 0}\frac{P_h-I}{h}$. The steady-state equations in this case are

$$\pi^\top P_h=\pi^\top$$
$$\pi\geq 0$$
$$\pi\cdot e=1$$

Feeding our expression for P_h into the first equation, we obtain

$$\pi^\top+\pi^\top Qh+o(h)=\pi^\top \Rightarrow \pi^\top Q=0$$

For our particular matrix, this give

$$-\lambda\pi_0+\mu\pi_1=0$$
$$\lambda\pi_{n-1}-(\mu+\lambda)\pi_n+\mu\pi_{n+1}=0$$

Letting $\rho=\lambda/\mu$ and assuming $\rho<1$, we obtain

$$\pi_0\rho=\pi_1$$
$$\rho\pi_{n+1}-(1+\rho)\pi_n+\pi_{n+1}=0$$

Solving this recurrence relation, we obtain

$$\pi_n=c\rho^n \qquad \sum\pi_n=1\Rightarrow c=1-\rho$$

So

$$\boxed{\pi_n = (1-\rho)\rho^n \qquad n \geq 0}$$

This expression passes a "sanity check" at $n = 0$; the probability the chain is empty is given by $\pi_0 = 1-\rho = 1-\frac{\lambda}{\mu}$, which we would expect to be the average amount of time the system is empty.

- We can use Little's Law to good effect; consider the following two examples; the first is trivial, the second less so
  - ***Consider the server as the "system"***

    Let S be the number of items in the server. The average sojourn time in the server is $\frac{1}{\mu}$. As such, by Little's Law

    $$\mathbb{E}_\pi(S) = \frac{\lambda}{\mu} = \rho$$

    Note, however, that $S = 1 \cdot \mathbb{1}_{\text{System is busy}} + 0 \cdot \mathbb{1}_{\text{System is idle}}$. As such, $\mathbb{E}_\pi(S) = \mathbb{P}(\text{System is busy}) = \rho$. The result above is therefore consistent with what we would expect.
  - ***Consider the queue as the "system"***

    Let Q be the number of items in the queue. The sojourn time in the queue is simply the waiting time, w_i. As such, by Little's Law

    $$\mathbb{E}_\pi(Q) = \lambda \mathbb{E}_\pi(w)$$

    (where w is the waiting time for any new item that joins the queue). Now, note that $Q(t) = (X(t)-1)^+$ (we must subtract any item currently in the server). As such

    $$\mathbb{E}_\pi(Q) = \sum_{n=1}^{\infty}(n-1)(1-\rho)\rho^n = \frac{\rho^2}{1-\rho}$$

    Therefore

    $$\mathbb{E}_\pi(w) = \frac{\rho}{\mu(1-\rho)}$$

- We know consider the more difficult problem of deriving the probability distribution of waiting times, $\mathbb{P}_\pi\left(w > x\right)$. Once again, recall w is the wait time experienced by a random job entering the queue.

$$\begin{aligned}\mathbb{P}_\pi\left(w > x\right) &= 0 \cdot \mathbb{P}\left(X = 0\right) + \sum_{k=1}^{\infty} \mathbb{P}_\pi\left(w > x \mid X = k\right)\mathbb{P}_\pi\left(X = k\right) \\ &= \sum_{k=1}^{\infty} \mathbb{P}_\pi\left(w > x \mid X = k\right)(1-\rho)\rho^k\end{aligned}$$

Now, note that since the exponential distribution is memoryless, we can write

$$\left(w \mid X = k\right) =_d \overbrace{S_k}^{\substack{\text{Remaining time} \\ \text{for job currently} \\ \text{in server}}} + \overbrace{\sum_{j=1}^{k-1} S_j}^{\substack{\text{Processing time} \\ \text{for the } k-1 \text{ other} \\ \text{jobs waiting}}} =_d \text{Erlang}(k, \mu)$$

(the Erlang distribution is the distribution of the sum of exponential variables; it is a special case of the Gamma distribution). As such

$$\begin{aligned}f_w(x) &= \sum_{k=1}^{\infty} \frac{e^{-\mu x}(\mu x)^{k-1}\mu}{(k-1)!}\rho^k(1-\rho) \\ &= e^{-\mu x}\mu(1-\rho)\rho\sum_{k=1}^{\infty}\frac{(\mu x)^{k-1}}{(k-1)!}\rho^{k-1} \\ &= e^{-\mu x}\mu(1-\rho)\rho\sum_{k=0}^{\infty}\frac{(\mu x\rho)^k}{k!} \\ &= \rho\mu(1-\rho)e^{-\mu(1-\rho)x}\end{aligned}$$

And so

$$w \sim \begin{cases}\exp\left(\mu(1-\rho)\right) & \text{with prob } \rho \\ 0 & 1-\rho\end{cases}$$

And

$$\mathbb{P}\left(w > x\right) = \rho e^{-\mu x(1-\rho)}$$

- ***The M/G/1 queue***
  - We now consider a slightly more general case in which the time between arrivals $\left\{\tau_n\right\}$ are exponential, but the processing times $\left\{S_n\right\}$ now follow a general distribution. This complicates matters slightly, because $X(t)$ – the amount of work in the system – is no sufficient enough to totally describe the state of the system at time t. We also require $R(t)$, the residual processing time of the work currently in service. (This was not the case when processing times followed a *memoryless* exponential distribution).

- To sidestep this complication, we will consider the following *embedded Markov chain* X_n. Let T_n be the time at which the n^{th} job concludes processing, and denote

$$X(T_{n^+}) \equiv X_n \equiv \begin{array}{c}\text{Number of jobs in the system} \\ \text{immediately after the } n^{\text{th}} \text{ job as departed}\end{array}$$

We then have

$$X_{n+1} = \left(X_n - 1\right)^+ + A_{n+1}$$

Where A_{n+1} is the number of arrivals during the processing time of the $(n+1)^{\text{th}}$ job. In this case, by assumption, the A_n are IID, and $\left(A_n \mid S_n = s\right) \sim \text{Po}\left(\lambda s\right)$, where λ is the rate of arrivals.

- Now, using the ergodic theorem for DTMCs (and assuming we have stability; ie: $\rho < 1$), we have

$$\mathbb{P}\left(X_n = j\right) \to_{n\to\infty} \pi(j)$$

Furthermore, from the $G/G/1$ case, we know that for each sample path, $X(t) = X_t^*$ exists, which implies that

$$\mathbb{P}\left(X(t) = j\right) \to_{t\to\infty} \hat{\pi}(j)$$

The challenge now, however, is to prove that $\pi = \hat{\pi}$.

- ***Theorem***: $\pi = \hat{\pi}$

Proof: Define the following two processes:

$$\begin{aligned} A_j(t) &= \text{Number of arrivals in } [0,t) \text{ that "find" the system in state } j \\ D_j(t) &= \text{Number of arrivals in } [0,t) \text{ that "leave" the system in state } j \end{aligned}$$

Also let $A(t)$ and $D(t)$ be the total number of arrivals and departures up to time t, respectively. We now derive or quote a number of seemingly unrelated facts and then combine them to prove our result

1. For any given sample path ω, it is clear that

$$\left|A_j(t) - D_j(t)\right| \le 1 \qquad \forall t \ge 0$$

This is because $A_j(t)$ is the number of "up-crossings" of $X(t)$ over the line $X(t) = j$ whereas $D_j(t)$ is the number of "down-crossings". Since the path is continuous, every up-crossing must be followed by a down-crossing, and so the difference in numbers will be at most 1.

2. Using the ergodic theorem of Markov Chains, we have

$$\lim_{t\to\infty} \frac{D_j(t)}{D(t)} = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{\{X_k = j\}} = \pi(j) \quad \text{a.s.}$$

(The first equality follows from the fact that X_k is precisely the state of the system after the k^{th} departure).

3. Using the bound in (1), we have that $D_j(t) - 1 \le A_j(t) \le D_j(t) + 1$, so

$$\overbrace{\frac{D(t)}{A(t)}}^{\to 1} \overbrace{\frac{D_j(t) - 1}{D(t)}}^{\to \pi(j) \text{ by } (2)} \le \frac{A_j(t)}{A(t)} \le \overbrace{\frac{D_j(t) + 1}{D(t)}}^{\to \pi(j) \text{ by } (2)} \cdot \overbrace{\frac{D(t)}{A(t)}}^{\to 1}$$

And so

$$\frac{A_j(t)}{A(t)} \to \pi(j) \text{ a.s.}$$

4. Letting t_k be the time of the k^{th} arrival, we have

$$\lim_{t\to\infty} \frac{A_j(t)}{A(t)} = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{\{X(t_k) = j\}}$$

5. A well-known property of queues with Poisson arrivals is PASTA (Poisson arrivals see time averages – see addendum at the end of this lecture).

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t \mathbb{1}_{\{X(s) = j\}} \, \mathrm{d}s = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{\{X(t_k) = j\}}$$

where t_k is the time of the k^{th} arrival to the system. Effectively, this states that in working out the average work in the queue, we don't need to sample at *every* time-step; it is enough to sample at arrivals.

6. By the ergodic theorem for Markov Chains

$$\hat{\pi}(j) =_{\text{.a.s}} \lim_{t\to\infty} \frac{1}{t}\int_0^t \mathbb{1}_{\{X(s) = j\}} \, \mathrm{d}s$$

7. Finally, combine all the above

$$\begin{aligned}\hat{\pi}(j) &\overbrace{=_{\text{a.s.}}}^{(6)} \lim_{t\to\infty} \frac{1}{t}\int_0^t \mathbb{1}_{\{X(s)=j\}}\,\mathrm{d}s \\ &\overbrace{=}^{(5)} \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{\{X(t_k)=j\}} \\ &\overbrace{=}^{(4)} \lim_{t\to\infty} \frac{A_j(t)}{A(t)} \\ &\overbrace{=_{\text{a.s.}}}^{(3)} \pi(j)\end{aligned}$$

Which proves the required result. ■

- ***Example***: We can use this result to derive the Pollaczek–Khinchine formula for the average waiting time in a $M/G/1$ queue. First, consider that by the above

$$\mathbb{E}_{\hat{\pi}}\left[X(t)\right] = \mathbb{E}_{\pi}\left[X_n\right]$$

Recall further that

$$X_{n+1} = \left(X_n - 1\right)^+ + A_{n+1} \qquad\qquad \left(A_n \mid S_n = s\right) \sim Po(\lambda s)$$

We can write this as follows

$$\begin{aligned} X_{n+1} &= (X_n - 1)\mathbb{1}_{\{X_n \geq 1\}} + A_{n+1} \\ &= X_n \mathbb{1}_{\{X_n \geq 1\}} - \mathbb{1}_{\{X_n \geq 1\}} + A_{n+1}\end{aligned}$$

And

Squaring this expression, we obtain (we simply denote $\mathbb{1} \equiv \mathbb{1}_{\{X_n \geq 1\}}$ to save space)

$$X_{n+1}^2 = X_n^2\mathbb{1} - 2X_n\mathbb{1} + 2X_nA_{n+1}\mathbb{1} + \mathbb{1} - 2A_{n+1}\mathbb{1} + A_{n+1}^2$$

As such

$$\mathbb{E}_{\pi}\left(X_{n+1}^2\right) = \mathbb{E}_{\pi}\left(X_n^2\mathbb{1} - 2X_n\mathbb{1} + 2X_nA_{n+1}\mathbb{1} + \mathbb{1} - 2A_{n+1}\mathbb{1} + A_{n+1}^2\right)$$

We can begin by simplifying the expression above by noting that $\mathbb{E}_{\pi}\left(X_n \mathbb{1}_{\{X_n \geq 1\}}\right) =$

And so

$$\begin{aligned}\mathbb{E}_{\pi} X_{n+1}^2 &= \mathbb{E}_{\pi} X_n^2 \mathbb{1}_{\{X_n \geq 1\}} + \mathbb{E}_{\pi}\mathbb{1}_{\{X_n \geq 1\}} + \mathbb{E}A_{n+1}^2 - 2X_{n+1}\mathbb{1} - 2A_{n+1}\mathbb{1} + 2X_nA_{n+1}\mathbb{1} \\ 0 &= \rho + \mathbb{E}\left(A_{n+1}^2\right) - 2\mathbb{E}_{\pi}(X) - 2\mathbb{E}(A_{n+1})\rho + 2\mathbb{E}(A_{n+1})\mathbb{E}_{\pi}(X)\end{aligned}$$

$\uparrow$ add indicator ≥ 0, same

Note that

$$\mathbb{E}(A_1) = \mathbb{E}\left[\mathbb{E}(A \mid S)\right] = \lambda\mathbb{E}(S) = \rho$$
$$\mathbb{V}\text{ar}(A_1) = \mathbb{E}\left[\mathbb{V}\text{ar}(A_1 \mid S)\right] + \mathbb{V}\text{ar}+([\text{E(A|S)}]))=\lambda\mathbb{E}(S) + \mathbb{V}\text{ar}(\lambda S) = \rho + \lambda^2\sigma_s^2$$

So

$$0 = \rho + \rho + \lambda^2\sigma_s^2 + \rho^2 - 2\mathbb{E}_\pi X - 2\rho^2 + 2\rho\mathbb{E}_\pi X$$
$$2\mathbb{E}_\pi X(1-\rho) = 2\rho + \lambda^2\sigma_s^2 + \rho^2$$
$$\mathbb{E}_\pi X = \frac{2\rho + \rho^2 + \lambda^2\sigma_s}{2(1-\rho)}$$
$$\mathbb{E}_\pi X = \frac{2\rho + \rho^2 + \lambda^2\mathbb{E}(S)^2\,\frac{\sigma_s^2}{(E[S])^2}}{2(1-\rho)}$$
$$\mathbb{E}_\pi X = \frac{2\rho + \rho^2(1 + CV^2)}{2(1-\rho)}$$

For M/M/1, CV = 1, so back to expression above.

$$\lambda\mathbb{E}_\pi w = \mathbb{E}_\pi(X-1)^+ = \mathbb{E}_\pi(X-1)\mathbb{1}_{\{X\geq 1\}} = \mathbb{E}_\pi X\mathbb{1}_{X\geq 1} - \mathbb{P}_\pi\left(X \geq 1\right) = \mathbb{E}_\pi(X) - \rho$$

Plugging in from the above

$$\lambda\mathbb{E}_\pi w = \frac{\rho^2(1+CV^2)+2\rho^2}{2(1-\rho)} = \frac{\rho}{\mu(1-\rho)} + \frac{\rho(1+CV^2)}{2\mu(1-\rho)}$$

All true only if we know the expectation of X^2 exists. Thus, out that $\mathbb{E}(S^3) < \infty \Leftarrow \mathbb{E}_\pi X^2 < \infty$, this is excessive because we took the simpler way to prove it. Really, all we need is the second moment. Easy approach to problem leads to an excessive condition.

Pillache-kinchin formula. The steady states being equal doesn't usually happen.

- ***Addendum on PASTA***
    - The principle of PASTA (Poisson Arrivals see Time Averages) states that for *any* stochastic process $X(t)$ over a state space $\mathcal{S}$, and for any $A \subseteq \mathcal{S}$

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\mathbb{1}_{\{X(t_k)\in A\}} = \lim_{t\to\infty}\frac{1}{t}\int_0^t \mathbb{1}_{\{X(s)\in A\}}\,\mathrm{d}s$$

provided the t_k form a Poisson process – in other words, $t_k - t_{k-1} \sim \exp(\lambda)$. This is really quite a surprising result! The RHS has full information and sees the system at *all* times. The LHS only sees it at a *selection* of times.

- For an intuitive proof, write $A(t)$ = cumulative number of samples up to time t. the statement above can then be written as

$$\lim_{t\to\infty} \frac{1}{A(t)} \int_0^t \mathbb{1}_{\{X(s)\in A\}} \, \mathrm{d}A(s) = \lim_{t\to\infty} \frac{1}{t} \int_0^t \mathbb{1}_{\{X(s)\in A\}} \, \mathrm{d}s$$

Now, write

$$\begin{aligned} M(t) &= \int_0^t \mathbb{1}_{\{X(s)\in A\}} \, \mathrm{d}A(s) - \lambda \int_0^t \mathbb{1}_{\{X(s)\in A\}} \, \mathrm{d}s \\ &= \int_0^t \mathbb{1}_{\{X(s)\in A\}} \, \mathrm{d}\tilde{A}(s) \\ &\approx \sum_{k=1}^n \mathbb{1}_{\{X_k\in A\}} \left(\tilde{A}(k+1) - \tilde{A}(k)\right) \end{aligned}$$

Where $\mathrm{d}\tilde{A}(s) = \mathrm{d}A(s) - \lambda \mathrm{d}s \Rightarrow \tilde{A}(t) = A(t) - \lambda t$. This is a martingale, because the increments $\mathrm{d}\tilde{A}(s)$ are independent (by the "independent increments" property of Poissons processes) and mean 0 (because of our normalization). Furthermore, the conditional increments are uniformly bounded (since they are counts). As such $\frac{M(t)}{t} \to 0$ a.s. In other words,

$$\lim_{t\to\infty} \frac{A(t)}{\lambda t} \frac{1}{A(t)} \int_0^t \mathbb{1}_{\{X(s)\in A\}} \, \mathrm{d}A(s) = \lim_{t\to\infty} \frac{1}{t} \int_0^t \mathbb{1}_{\{X(s)\in A\}} \, \mathrm{d}s$$

Finally, use $\lim_{t\to\infty} \frac{A(t)}{t} = \lambda$ to conclude that the first term converges to 1. This gives the required result.

LECTURE 10 – 30TH MARCH 2011

Renewal & Regenerative Process

- ***Renewal processes***
  - Let $\{X_n\}$ be IID, with $\mathbb{E}(X_1) = \mu < \infty$ and $\mathbb{P}\left(X_1 = 0\right) < 1$. Let $S_n = \sum_{i=1}^n X_i$, with $S_0 = 0$. Let $N(t) = \sup\left\{n \geq 1 : S_n \leq t\right\}$. $\left\{N(t) : t \geq 0\right\}$ is called a *renewal process*.
  - ***Definition (renewal function)***: The *renewal function* is defined as $m(t) = \mathbb{E}[N(t)]$.

- ***Example***: Let $X_1 \sim \exp(1/\mu)$. $\{N(t) : t \geq 0\}$ is then a Poisson process. Consider, incidentally, that in a Poisson process, the following two facts are true. Much of our work in this section will be concerned with generalizing these results to general renewal processes:
  - $\mathbb{E}(N(t)) = m(t) = \mu t = \mathbb{E}(X_1)t$ (generalizes to the *elementary renewal theorem*).
  - $m'(t) = \mu = \mathbb{E}(X_1)$ (generalizes to the *key renewal theorem*). □

- ***Proposition***: $m(t) = \sum_{n=1}^{\infty} F_n(t)$, where $F_n(t) = \mathbb{P}\left(S_n \leq t\right)$.

  Proof: Note that

$$N(t) = \sum_{n=1}^{\infty} \mathbb{1}_{\{S_n \leq t\}}$$
$$\mathbb{E}[N(t)] = \sum_{n=1}^{\infty} \mathbb{P}\left(S_n \leq t\right)$$

  Where, in the last line, we used Fubini I. ■

  Remark: The CDF of S_n is the n-fold convolution of the CDF of each individual variable X.

$$F_n(\cdot) = \underbrace{F_1(\cdot) * F_1(\cdot) * \cdots * F_1(\cdot)}_{n \text{ fold}}$$

  For example,

$$F_2(t) = \int F(t-n) \, \mathrm{d}F(n)$$

- ***Laplace Transforms & Co.***
  - The Laplace Transform of a function of a distribution with CDF $F(x)$ is defined as

$$\hat{F}(s) = \int_0^{\infty} e^{-sx} \, \mathrm{d}F(x) = \int_0^{\infty} e^{-sx} f(x) \, \mathrm{d}x$$

    A few important results:
    - Laplace transform of convolutions. In CDF form, if $F = A * B$, then
$$\hat{F}(s) = \hat{A}(s)\hat{B}(s)$$
    - Note that if f is a density function, then
$$\left|\hat{F}(s)\right| = \left|\int_0^{\infty} e^{-sx} \, \mathrm{d}F(x)\right| < 1$$
  - Together, the points above imply that

$$\hat{m}(s) = \sum_{n=1}^{\infty} \hat{F}(s)^n = \frac{\hat{F}(s)}{1 - \hat{F}(s)}$$

Similarly, re-arranging

$$\hat{F}(s) = \frac{\hat{m}(s)}{1 + \hat{m}(s)}$$

As such, there is a one-to-one correspondence between F and m.

- ***Theorem***: If $b(t)$ is bounded on any interval, then the solution to the following *renewal equation*

$$a = b + a * F$$
$$a(t) = b(t) + \int_0^\infty a(t-s) \, \mathrm{d}F(s)$$

is

$$a = b + b * m$$
$$a(t) = b(t) + \int_{s=0}^{t} b(t-s) \, \mathrm{d}m(s)$$

Verification: Taking the Laplace transform of the first equation

$$\hat{a}(s) = \hat{b}(s) + \hat{a}(s)\hat{F}(s)$$

As such

$$\hat{a}(s) = \frac{\hat{b}(s)}{1 - \hat{F}(s)} = \hat{b}(s) + \frac{\hat{b}(s)\hat{F}(s)}{1 - \hat{F}(s)} = \hat{b}(s) + \hat{b}(s)\hat{m}(s)$$

Taking an inverse Laplace transform of the proposed solution, we do indeed find this equation is satisfied. ■

- ***Example***: Consider that

$$\mathbb{E}\left[N(t) \mid X_1 = x\right] = \begin{cases} 0 & x > t \\ 1 + \mathbb{E}\left[N(t) - N(x) \mid X_1 = x\right] & x \le t \end{cases}$$
$$= \begin{cases} 0 & x > t \\ 1 + m(t-x) & x \le t \end{cases}$$

However,

$$\begin{aligned} m(t) &= \mathbb{E}\big[N(t)\big] \\ &= \int_0^\infty \mathbb{E}\big[N(t) \mid X_1 = k\big] \, \mathrm{d}F(x) \\ &= \int_0^t \big[1 + m(t-x)\big] \, \mathrm{d}F(x) \\ &= F(t) + \int_0^t m(t-x) \, \mathrm{d}F(x) \\ &= F(t) + (m * F)(t) \end{aligned}$$

This equation is in the form of a renewal equation, with $b = F$. The solution is therefore

$$m(t) = F(t) + \int_{s=0}^t F(t-s) \, \mathrm{d}m(s) \qquad \square$$

- ***Example***: Let us consider the example $X \sim \exp(1/\mu)$ and $m(t) = t\mu \Rightarrow \mathrm{d}m(t) = \mu \mathrm{d}t$. Thus, the solution to the standard renewal equation is

$$a(t) = b(t) + \mu \int_{s=0}^t b(t-s) \, \mathrm{d}s = b(t) + \mu \int_{s=0}^t b(s) \, \mathrm{d}s$$

For the integral to be finite, we need $b(t) \to 0$. As such, we might expect that

$$a(t) \to \mu \int_{s=0}^t b(s) \, \mathrm{d}s \qquad \text{as } t \to \infty$$

It turns out this last conclusion holds generally, not just for Poisson processes.

- ## *Some Theorems*

- ***Proposition (SLLN for renewal processes)***

$$\frac{N(t)}{t} \to \frac{1}{\mathbb{E}(X_1)} \text{ a.s. as } t \to \infty$$

Proof: Consider that $S_{N(t)} \leq t \leq S_{N(t)+1}$. As such

$$\frac{S_{N(t)}}{N(t)} \leq \frac{t}{N(t)} \leq \frac{S_{N(t)+1}}{N(t)+1} \underbrace{\frac{N(t)+1}{N(t)}}_{\to 1 \text{ a.s.}}$$

However, by the SLLN, $\frac{S_n}{n} \to \mathbb{E}(X_1)$. Feeding this into the above proves our theorem. ■

- ***Proposition (Elementary/Baby Renewal Theorem)***:

$$\frac{m(t)}{t} = \frac{\mathbb{E}\big[N(t)\big]}{t} \to \frac{1}{\mathbb{E}(X_1)} \qquad \text{as } t \to \infty$$

Proof: See homework 4.

- ***Proposition (CLT for renewal processes)***: Suppose that $\mathbb{V}\mathrm{ar}[X_1] = \sigma^2 < \infty$

$$\sqrt{t}\left(\frac{N(t)}{t}-\frac{1}{\mathbb{E}(X_1)}\right)\Rightarrow\sigma\frac{1}{\left[\mathbb{E}(X_1)\right]^{3/2}}N(0,1)$$

Or in other words

$$N(t)\approx\frac{t}{\mathbb{E}(X_1)}+\sqrt{t}\frac{\sigma}{\left[\mathbb{E}(X_1)\right]^{3/2}}N(0,1)$$

Proof: Throughout this proof, we will write $\mathbb{E}(X_1)=\mu$.

1. ***Step 1***: We first show that

$$\sqrt{t}\left(\frac{N(t)}{S_{N(t)}}-\frac{1}{\mu}\right)\Rightarrow\sigma\frac{1}{\mu^{3/2}}N(0,1)$$

We do this as follows:

$$\begin{aligned}\sqrt{t}\left(\frac{N(t)}{S_{N(t)}}-\frac{1}{\mu}\right)&=\frac{\sqrt{t}}{S_{N(t)}/N(t)}\left(1-\frac{S_{N(t)}}{\mu N(t)}\right)\\&=\frac{\sqrt{t}}{\mu}\frac{1}{S_{N(t)}/N(t)}\left(\mu-\frac{S_{N(t)}}{N(t)}\right)\\&=\underbrace{\sqrt{\frac{t}{N(t)}}}_{\to\sqrt{\mu}}\frac{1}{\mu}\underbrace{\frac{1}{S_{N(t)}/N(t)}}_{\to\mu}\sqrt{N(t)}\left(\mu-\frac{S_{N(t)}}{N(t)}\right)\end{aligned}$$

Now, consider the last term – since $N(t)\to\infty$, it is a subsequence of $\sqrt{n}\left(\frac{S_n}{n}-\mu\right)$. This latter expression does tend to $\sigma N(0,1)$ in distribution, by the standard CLT. The question is whether the weak convergence also holds for the subsequence. To answer this question, we will require a side-lemma.

Lemma (Ansombe's Lemma/random time change lemma): If $\{Z_n\}$ are IID with mean 0 and $\mathbb{E}(Z_1^2)=\sigma^2<\infty$ and $\{N(t):t\geq 0\}$ is an integer valued process such that $\frac{N(t)}{t}\Rightarrow\beta\in(0,\infty)$, then

$$\frac{1}{\sqrt{N(t)}}\sum_{i=1}^{N(t)}Z_1\Rightarrow\sigma N(0,1)$$

In other words, we require our random sequence to converge at linear linearly to infinity for convergence in distribution to hold. Intuitively, this is because we need variances to accumulate fast enough for the CLT to be valid.

In this case, note that

$$\frac{N(t)}{t} \to \frac{1}{\mathbb{E}(X_1)} \text{ a.s.}$$

This implies that

$$\frac{N(t)}{t} \Rightarrow \frac{1}{\mathbb{E}(X_1)} \left(= \beta\right)$$

Hence, the Ansombe's Lemma applies. We can then use the converging together Lemma on the expression quoted above to obtained the desired result for this step.

Step 2: We now use step 1 to prove our result. Consider that

$$\frac{S_{N(t)}}{N(t)} \le \frac{t}{N(t)} \le \frac{S_{N(t)+1}}{N(t)+1} \cdot \underbrace{\frac{N(t)+1}{N(t)}}_{\to 1 \text{ a.s.}}$$

Both sides of the "sandwich" are in the form used in step 1. Therefore, this immediately leads to the desired result. ■

- ***Definition (Lattice/Arithmetic distribution)***: A distribution F is said to be of Lattice-type if there exists an $h > 0$ such that F is supported on

$$\left\{X_n\right\} = \left\{nh : n \in \mathbb{Z}\right\}$$

For example, the Poisson distribution is lattice with $h = 1$.

- ***Theorem (Blackwell's Theorem)***: If F is *non-lattice*, then

$$m(t+a) - m(t) \to \frac{1}{\mathbb{E}(X_1)} \cdot a \qquad \text{as } t \to \infty$$

For all $a > 0$.

Remark: This result is *not* implied by the fact $\frac{m(t)}{t} \to \frac{1}{\mathbb{E}(X_1)}$, because this theorem concerns *increments* in the renewal function m.

- ***Definition***: A function $f : \mathbb{R}_+ \to \mathbb{R}_+$ is *directly Riemann integrable (dRi)* if

- $\int_0^\infty \overline{f}_n(x)\ \mathrm{d}x < \infty$
- $\int_0^\infty \overline{f}_n(x)\ \mathrm{d}x - \int_0^\infty \underline{f}_n(x)\ \mathrm{d}x \to_{h\to 0} 0$

Where

$$\overline{f}_n(x) = \sum_{k=0}^\infty \sup\left\{f(x) : kn \le x \le (k+1)n\right\} \mathbb{1}_{\{x\in[kh,(k+1)h\}}$$
$$\underline{f}_n(x) = \sum_{k=0}^\infty \inf\left\{f(x) : kn \le x \le (k+1)n\right\} \mathbb{1}_{\{x\in[kh,(k+1)h\}}$$

Direct Riemann Integrability can be thought of as an extension of Riemann Integrability over an infinite line, as opposed to over an integral.

Remark: If $f : \mathbb{R}_+ \to \mathbb{R}$, we simply require both f^+ and f^- to be dRi for f to be dRi.

Remark: *Any* of the following conditions are sufficient for $f : \mathbb{R}_+ \to \mathbb{R}$ to be dRi

- f is continuous with compact support.
- f is bounded and continuous, and

  $$\int_0^\infty \overline{f}_h(x)\ \mathrm{d}x < \infty$$

  for some $h > 0$.
- f is non-increasing and $\int_0^\infty f(x)\ \mathrm{d}x < \infty$

- ***Theorem (Key Renewal Theorem)***: Providing the usual assumptions on F holds (IID increments with finite mean that are not masses at 0) and that the renewal process is non-lattice, then for any $b : \mathbb{R}_+ \to \mathbb{R}$ that is dRi

  $$a(t) \to \frac{1}{\mu}\int_0^\infty b(s)\ \mathrm{d}s \qquad\qquad t \to \infty$$

  Where a is a solution to renewal equation $a = b + a * F$.

LECTURE 11 – 6TH APRIL 2011

- ***Regenerative processes***
  - A regenerative process is a stochastic process with time points at which, from a probabilistic point of view, the process restarts itself. A good example is CTMC (for example, corresponding to an $M/M/1$ queue). We consider that the process "resets" each time the queue empties.

- ***Definition***: Let $\left(X(t):t\geq 0\right)$ be a stochastic process with a sample path that is right-continuous with left limits (RCLL or CADLAG). Without loss of generality, we let the renewal occur at $X(t)=0$. Now, define the following quantities:
  - $\tau(n+1)=\inf\left\{t\geq\tau(n):X(t)=0,X(t_{-})\neq 0\right\}$, the time of the n^{th} renewal.
  - $\tau_{n+1}=\tau(n+1)-\tau(n)$, the inter-renewal time.
  - $\tilde{X}_n(t)=\begin{cases}X\left(\tau(n-1)+t\right) & t\in[0,\tau_n]\\ \Delta & t>\tau_n\end{cases}$, where Δ is outside the state space of X.

  Then X is said to be *regenerative* if
  - $\tau_n<\infty$ a.s. $\forall n$
  - $\tau(n)\to\infty$ as $n\to\infty$
  - $\tilde{X}_0,\tilde{X}_1,\cdots$ are independent
  - $\tilde{X}_1,\tilde{X}_2,\cdots$ are identically distributed (the first renewal might have a different distribution if the system doesn't start "empty").
- ***Example (CTMC)***: Let $\left(X(t):t\geq 0\right)$ be a CTMC with state space $\mathcal{S}=\mathbb{Z}$. We can then define a renewal process using *any* state $x\in\mathcal{S}$ as our "renewal state". Our only requirement is that x be *at least* null recurrent; we must return to the state infinitely often. □
- ***Example (DTMC)***: Let $\left(X_n:n\geq 0\right)$ be a DTMC on $\mathcal{S}=\mathbb{Z}$. We can then define a renewal process using *any* state $x\in\mathcal{S}$, provided the chain is irreducible and recurrent. □
- ***Definition (recurrence)***: A regenerative process is *positive-recurrent* if $\mathbb{E}(\tau_1)<\infty$ and *null-recurrent* otherwise.
- ***Proposition (SLLN for regenerative processes)***: Let $\left(X(t):t\geq 0\right)$ be a regenerative process over a state space $\mathcal{S}$ with $\mathbb{E}(\tau_1)<\infty$, and let $f:\mathcal{S}\to\mathbb{R}$ be such that

$$\mathbb{E}\left(\int_{\tau(0)}^{\tau(1)}\left|f\left(X(s)\right)\right|\,\mathrm{d}s\right)<\infty$$

Then

$$\frac{1}{t}\int_0^t f\left(X(s)\right)\,\mathrm{d}s \to \frac{\mathbb{E}\left(\int_{\tau(0)}^{\tau(1)} f\left(X(s)\right)\,\mathrm{d}s\right)}{\mathbb{E}(\tau_1)} \text{ a.s. as } t\to\infty$$

Proof: Let

$$Y_k(f) = \int_{\tau(k-1)}^{\tau(k)} f\left(X(s)\right)\,\mathrm{d}s \qquad \text{and} \qquad N(t) = \sup\left\{n \geq 0 : \tau_n \leq t\right\}$$

This means that we can write

$$c(t) = \int_0^t f\left(X(s)\right)\,\mathrm{d}s = \sum_{k=0}^{N(t)} Y_k(f) + \int_{\tau\left(N(t)\right)}^t f\left(X(s)\right)\,\mathrm{d}s$$

As such

$$\frac{c(t)}{t} = \frac{Y_0(f)}{t} + \frac{1}{t}\sum_{k=1}^{N(t)} Y_k(f) + \frac{1}{t}\int_{\tau\left(N(t)\right)}^t f\left(X(s)\right)\,\mathrm{d}s$$

By the assumption in the theorem, $Y_0(f)$ is finite, and so the first term vanishes in the limit. Now

$$\frac{1}{t}\sum_{k=1}^{N(t)} Y_k(f) = \frac{N(t)}{t}\frac{1}{N(t)}\sum_{k=1}^{N(t)} Y_k(f)$$

Note, however, that the $Y_i(f)$ are IID. Furthermore,

$$\mathbb{E}\left|Y_k(f)\right| = \mathbb{E}\left|\int_{\tau(k-1)}^{\tau(k)} f\left(X(s)\right)\,\mathrm{d}s\right| = \mathbb{E}\left|\int_{\tau(0)}^{\tau(1)} f\left(X(s)\right)\,\mathrm{d}s\right| \leq \mathbb{E}\left(\int_{\tau(0)}^{\tau(1)} \left|f\left(X(s)\right)\right|\,\mathrm{d}s\right) < \infty$$

by assumption, and so we can use the SLLN (since $N(t)\to\infty$), and the SLLN for renewal processes to write the above as

$$\frac{1}{t}\sum_{k=1}^{N(t)} Y_k(f) = \frac{1}{\mathbb{E}(\tau_1)}\mathbb{E}\left(Y_1(f)\right)$$

Precisely as required. All we now need to do is to show that the last term vanishes.

$$\begin{aligned}\left|\frac{1}{t}\int_{\tau\left(N(t)\right)}^t f\left(X(s)\right)\,\mathrm{d}s\right| &\leq \frac{1}{t}\int_{\tau\left(N(t)\right)}^t \left|f\left(X(s)\right)\right|\,\mathrm{d}s \\ &\leq \frac{N(t)}{t}\frac{1}{N(t)}\int_{\tau\left(N(t)\right)}^t \left|f\left(X(s)\right)\right|\,\mathrm{d}s\end{aligned}$$

We now use the following Lemma.

Lemma: If $\{a_n\}$ is a real-valued sequence such that $\frac{1}{n}\sum_{i=1}^{n} a_i \to a_\infty$, then $\frac{1}{n}\max_{1\le i\le n}\{a_i\} \to 0$ as $n \to \infty$.

Notice that $\frac{1}{N(t)}\sum_{k=1}^{N(t)} Y_k(f) \to_{\text{a.s.}} \mathbb{E}\left(Y_1(f)\right)$, since $N(t)\to\infty$. As such, $\frac{1}{N(t)}\max_{1\le K\le N(t)}\left\{Y_k(f)\right\} \to 0$. As such

$$\left|\frac{1}{t}\int_{\tau(N(t))}^{t} f\left(X(s)\right)\,\mathrm{d}s\right| \le \frac{N(t)}{t}\cdot\frac{\max_{1\le K\le N(t)+1}\left\{Y_k(f)\right\}}{N(t)} \to \frac{1}{\mathbb{E}(\tau_1)}\cdot 0 \text{ a.s.}$$

As expected. ■

- We have just proved that $\frac{1}{t}\int_0^t f\left(X(s)\right)\,\mathrm{d}s \to_{\text{a.s.}} \frac{\mathbb{E}(Y_1(f))}{\mathbb{E}(\tau_1)}$. This suggests that $\mathbb{E}_\pi\left[f\left(X(s)\right)\right] = \frac{\mathbb{E}(Y_1(f))}{\mathbb{E}(\tau_1)}$.
- ***Proposition (CLT for regenerative processes)***: Let $\left(X(t): t\ge 0\right)$ be a regenerative process on a state space $\mathcal{S}$ with $\mathbb{E}(\tau_1^2)<\infty$. Let $f:\mathcal{S}\to\mathbb{R}$ be a function satisfying $\mathbb{E}\left(Y_1(|f|)^2\right)<\infty$. Then

$$\sqrt{t}\left(\frac{\int_0^t f\left(X(s)\right)\,\mathrm{d}s}{t} - \frac{\mathbb{E}\left(Y_1(t)\right)}{\mathbb{E}(\tau_1)}\right) \Rightarrow \sigma N\left(0,1\right)$$

With $\sigma^2 = \mathbb{E}\left(Y_1^2(f_c)\right)/\,\mathbb{E}(\tau_1)$ where $f_c(\cdot) = f(\cdot) - \frac{\mathbb{E}(Y_1(f))}{\mathbb{E}(\tau_1)}$.

- ***Theorem***: Let $\left(X(t): t\ge 0\right)$ be a positive recurrent regenerative process on a state space $S\subseteq\mathbb{R}^d$. Suppose that either of the following conditions hold
  - $F(x) = \mathbb{P}\left(\tau_1 = x\right)$ has a density and there is a function $h:\mathbb{R}^d\to\mathbb{R}$ that is bounded
  - $F(x) = \mathbb{P}\left(\tau_1 = x\right)$ is non-lattice and there is a function h that is bounded and continuous

  Then

$$\mathbb{E}\left[h(X(t))\right] \to \frac{\mathbb{E}\left[\int_{\tau(0)}^{\tau(1)} h(X(s))\,\mathrm{d}s\right]}{\mathbb{E}\left[\tau_1\right]} \qquad \text{as } t\to\infty$$

  Then

$$X(t) \Rightarrow X(\infty) \qquad \text{as } t\to\infty$$

and

$$\mathbb{P}\Big(X(t) \in A\Big) = \frac{\mathbb{E}\left[\int_{\tau(0)}^{\tau(1)} \mathbb{1}_{\{X(s)\in A\}} \ \mathrm{d}s\right]}{\mathbb{E}\left[\tau_1\right]}$$

Remark: If $X(0) =_d X(\infty)$, then $\Big(X(t) : t \geq 0\Big)$ is stationary.

Proof: Fix $t > 0$, and suppose $\tau(0) = 0$ (without loss of generality).

$$\begin{aligned}
a(t) &= \mathbb{E}\Big[h(X(t))\Big] \\
&= \mathbb{E}\Big[h(X(t)), \tau_1 > t\Big] + \mathbb{E}\Big[h(X(t)), \tau_1 \leq t\Big] \\
&= b(t) + \int_0^t \mathbb{E}\Big[h(X(t)) \mid \tau_1 = s\Big] \ \mathrm{d}F(s) \\
&= b(t) + \int_0^t \mathbb{E}\Big[h(X(\tau(1) + t - s))\Big] \ \mathrm{d}F(s) \\
&= b(t) + \int_0^t \mathbb{E}\Big[h(X(\tau(0) + t - s))\Big] \ \mathrm{d}F(s)
\end{aligned}$$

(Where, in the last step, we have used the fact that every cycle has the same distribution). As such

$$a(t) = b(t) + \int_0^t a(t-s) \ \mathrm{d}F(s)$$
$$a = b + a * F$$

Now, consider that

$$b(t) = \mathbb{E}\Big[h(X(t)), \tau_1 > t\Big]$$
$$\Big|b(t)\Big| \leq \mathbb{E}\Big[\Big|h(X(t))\Big|, \tau_1 > t\Big] \leq \Big\|h\Big\|_\infty \mathbb{P}\Big(\tau_1 > t\Big) = g(t)$$

Where $g(t)$ is bounded and continuous and non-increasing and $\int_0^\infty g(t) \ \mathrm{d}t = \Big\|h\Big\|_\infty \mathbb{E}(\tau_1) < \infty$ finite. As such, by the conditions for direct Reimann integrability, we have that g is dRi. Now, if $\Big|b\Big| \leq g$ and g is dRi, then b is dRi. And since F is non-lattice, all the conditions of the key renewal theorem hold, and therefore

$$\begin{aligned}
a(t) \to_{t\to\infty} \ & \frac{1}{\mathbb{E}(\tau_1)} \int_0^\infty b(s) \ \mathrm{d}s \\
&= \frac{1}{\mathbb{E}(\tau_1)} \int_0^\infty \mathbb{E}\Big[h(X(s)), \tau_1 > s\Big] \ \mathrm{d}s \\
&= \frac{1}{\mathbb{E}(\tau_1)} \int_0^\infty \mathbb{E}\Big[h(X(s))\mathbb{1}_{\{\tau_1 > s\}}\Big] \ \mathrm{d}s \\
&= \frac{1}{\mathbb{E}(\tau_1)} \int_0^{\tau_1} h(X(s)) \ \mathrm{d}s
\end{aligned}$$

Where the last step follows by Fubini, since h is bounded. Now, since h is *any* bounded function, the above implies that $X(t) \Rightarrow X(\infty)$, and finally, taking $h_x(\cdot) = \mathbb{1}_{\{\cdot = x\}}$, we recover the last required statement.

- ***Example***: Let $\left(X(t) : t \geq 0\right)$ be a positive recurrent regenerative process. Fix a set $A \in S$, and let

$$T(A) = \inf\left\{t \geq 0 : X(t) \in A\right\}$$

Again, for simplicity, assume $\tau(0) = 0$. We are now interested in the expression $\mathbb{P}\left(T(A) > t\right) = ?$, especially for t large. For $t > 0$, note that

$$\begin{aligned} a(t) &= \mathbb{P}\left(T(A) > t\right) \\ &= \mathbb{P}\left(T(A) > t, \tau_1 > t\right) + \mathbb{P}\left(T(A) > t, \tau_1 \leq t\right) \\ &= \mathbb{P}\left(\left[T(A) \wedge \tau_1\right] > t\right) + \int_0^t \mathbb{P}\left(T(A) > t, \tau_1 = s\right) \, \mathrm{d}s \\ &= \mathbb{P}\left(\left[T(A) \wedge \tau_1\right] > t\right) + \int_0^t \mathbb{P}\left(T(A) > t, T(A) > \tau_1 \mid \tau_1 = s\right) \, \mathrm{dF}s \\ &= \mathbb{P}\left(\left[T(A) \wedge \tau_1\right] > t\right) + \int_0^t \mathbb{P}\left(\tilde{T}(A) > t - s, T(A) > \tau_1 \mid \tau_1 = s\right) \, \mathrm{dF}s \end{aligned}$$

Where $\tilde{T}(A) = \inf\left\{t > \tau(1), X(t) \in A\right\} = \inf\left\{t \geq 0, X(t + \tau(1)) \in A\right\}$. Now

$$\begin{aligned} a(t) &= \mathbb{P}\left(\left[T(A) \wedge \tau_1\right] > t\right) + \int_0^t \mathbb{P}\left(T(A) > t - s\right)\mathbb{P}\left(T(A) > \tau_1 \mid \tau_1 = s\right) \, \mathrm{dF}s \\ &= \mathbb{P}\left(\left[T(A) \wedge \tau_1\right] > t\right) \\ &\qquad + \int_0^t \mathbb{P}\left(T(A) > t - s\right)\mathbb{P}\left(\tau_1 = s \mid T(A) > \tau_1\right)\mathbb{P}\left(T(A) > \tau_1\right) \, \mathrm{dF}s \\ &= b(t) + \int_0^t a(t - s)\beta \, \mathrm{d}\tilde{\mathrm{F}}s \end{aligned}$$

Where

$$\tilde{F}(s) = \mathbb{P}\left(\tau_1 \leq s \mid T(A) > \tau_1\right)$$

So

$$a = b + \beta\left(a * F\right)$$

For fixed $\lambda \in \mathbb{R}$,

$$\begin{aligned} a(t)e^{\lambda t} &= b(t)e^{\lambda t} + \beta e^{\lambda t}\int_0^t a(t - s) \, \mathrm{d}\tilde{F}(s) \\ a_\lambda(t) &= b_\lambda(t) + \int_0^t \beta e^{\lambda(t-s)} a(t - s)\beta e^{\lambda s} \, \mathrm{d}\tilde{F}(s) \\ a_\lambda(t) &= b_\lambda(t) + \int_0^t a_\lambda(t - s) \, \mathrm{d}F_\lambda(s) \qquad \mathrm{d}F_\lambda(s) = \beta e^{\lambda s}\mathrm{d}\tilde{F}(s) \end{aligned}$$

Suppose $\exists\lambda$ s.t. F_λ is a bona-fide distribution. Then

$$F_\lambda(\infty) = 1 = \beta\int_0^\infty e^{\lambda s} \, \mathrm{d}\tilde{F}(s)$$

$$\beta^{-1} = \int_0^\infty e^{\lambda s}\ \mathrm{d}\tilde{F}(s) = \tilde{\mathbb{E}}\left(e^{\lambda Z}\right)$$

Assume there is such a solution λ^* [can show using hyeristic argument]. By appealing to the key renewal theorem

$$a_\lambda(t) \to \frac{1}{\mathbb{E}_\lambda\left(\tau_1\right)}\int_0^\infty e^{\lambda s}\mathbb{P}\left(\left[T(A) \wedge \tau_1\right] > s\right)\ \mathrm{d}s = \eta \qquad t \to \infty$$

$$\mathbb{E}_\lambda \tau_1 = \int_0^\infty x\ \mathrm{d}F_\lambda(x)$$

And so

$$a(t) = \mathbb{P}\left\{T(A) > t\right\} \sim \eta e^{-\lambda t}$$

So the probability has an exponential type tail. □

Index

Bold numbers indicate definitions. *Italicized* numbers indicate examples.