Qualifying Exam Prep Session #3: Modes of Convergence and Stochastic Processes

1 Convergence (BST 230 Module 12/13)

1.1 Deterministic Convergence

1.1.1 Convergence and limits

Let $x_1, x_2, \ldots$ be a sequence of real numbers. We say that the sequence (x_n) converges to $x \in \mathbb{R}$ if

$$\forall \varepsilon > 0 : \exists N \in \mathbb{N} \text{ s.t. } n \geqslant N \implies |x_n - x| < \varepsilon.$$

In other words, for any level of "closeness" ε, there is some point in the sequence after which the sequence is always ε-close to x.

We denote convergence in a number of different ways:

$$\begin{aligned} x_n &\to x \\ x_n &\to x \text{ as } n \to \infty \\ x_n &\overset{n\to\infty}{\to} x \\ \lim_{n\to\infty} x_n &= x. \end{aligned}$$

As is often the case when we move from the finite to the infinite, our ideas no longer work quite as we might hope. The idea that $|x_n - \infty| < \varepsilon$ is not well-defined, but it will still be useful for us to reason about convergence to $\pm\infty$. We say that the sequence (x_n) converges to ∞ if

$$\forall c \in \mathbb{R} : \exists N \in \mathbb{N} \text{ s.t. } n \geqslant N \implies x_n > c.$$

Likewise, $x_n \to -\infty$ if $n \geqslant N \implies x_n < -c$.

1.1.2 The limit does not exist!

Limits are not guaranteed to exist; that is, it may be that a sequence (x_n) neither eventually settles around a real number x, nor explodes to $\pm\infty$. We'd still like to be able to reason about asymptotic behavior of the sequence in this context, so we generalize the notion of limits to the *liminf* and *limsup*, which are guaranteed to exist.

The liminf and limsup can be thought of as asymptotic lower/upper bounds on the sequence (x_n). In particular, let $\bar{R} = R \cup \{-\infty, \infty\}$. Then

$$\begin{aligned} \liminf_{n\to\infty} x_n &= \inf\{l \in \bar{R} : \forall \varepsilon > 0, \exists N \in N \text{ s.t. } n \geqslant N \implies x_n > l - \varepsilon\} \\ \limsup_{n\to\infty} x_n &= \inf\{u \in \bar{R} : \forall \varepsilon > 0, \exists N \in N \text{ s.t. } n \geqslant N \implies x_n < u + \varepsilon\}. \end{aligned}$$

It is always the case that $\liminf x_n \leqslant \limsup x_n$ and that $\lim x_n$, if it exists, is sandwiched between them. Thus, one way to show that $\lim x_n = x$ is to show that $\liminf x_n = \limsup x_n = x$.

1.2 Convergence of Random Variables

1.2.1 (Almost) Sure Convergence

The ideas above from deterministic convergence can be generalized to "probabilistic convergence", i.e. the convergence of random variables. To show how this happens, we'll start with a notion of convergence of r.v. which (I believe) was not covered in BST 230.

Suppose we have a sequence of random variables $(X_n) = \{X_n\}_{n\in\mathbb{N}}$ each defined over a shared probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 1.1 (Sure (Pointwise) Convergence). *The sequence (X_n) converges surely to a r.v. X if*

$$\forall \omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega).$$

Unpacking this a little, recall that we define random variables as maps from the sample space Ω to $\mathbb{R}$. The "randomness" comes from the probability function $\mathbb{P}$ that assigns probabilities to elements of $\mathcal{F}$ (i.e. subsets of Ω).

Sure convergence is not very interesting from a probabilistic perspective, because it doesn't involve $(\mathcal{F}, \mathbb{P})$ at all, but it's an instructive place to start. From here, we can imagine a sequence which (X_n) which isn't guaranteed to converge for every ω, but for which the "bad" ω never occur under our probability model.

Definition 1.2 (Almost sure convergence). *We say that X_n converges to X almost surely (that is, $X_n \overset{a.s.}{\to} X$) if for*

$$A = \{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}$$

we have $\mathbb{P}(A) = 1$.

Like sure convergence, it is very natural to think of almost sure convergence in terms of deterministic convergence for individual ω values. Under almost sure convergence, there is allowed to be a set of ω on which $X_n(\omega) \not\to X(\omega)$, but this set must have probability 0.

Fact 1.3. *Almost sure convergence (alternate definition) $X_n \overset{a.s.}{\to} X$ if and only if, for all $\varepsilon > 0$:*

$$\mathbb{P}\left(\limsup_{n\to\infty} |X_n - X| < \varepsilon\right) = 1.$$

To show almost sure convergence, we typically use an (asymptotic) zero-one law, which give the conditions under which (asymptotically) a certain events happens either with probability 0 or 1. The Borel-Cantelli Lemmas are examples of zero-one laws.

Lemma 1.4 (First Borel-Cantelli). *Let (E_n) be an infinite sequence of events in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then,*

$$\sum_{n=1}^{\infty} \mathbb{P}(E_n) < \infty \implies \mathbb{P}\left(\limsup_{n\to\infty} E_n\right) = 0.$$

Note the connection here with Fact 1.3; one can show $X_n \overset{a.s.}{\to} X$ by showing that $\sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| > \varepsilon) < \infty$ for all $\varepsilon > 0$.

Lemma 1.5 (Second Borel-Cantelli). *Let (E_n) be an infinite sequence of independent events in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then,*

$$\sum_{n=1}^{\infty} \mathbb{P}(E_n) = \infty \implies \mathbb{P}\left(\limsup_{n\to\infty} E_n\right) = 1.$$

Perhaps the most famous use of an almost sure convergence result is in the *strong law of large numbers.*

Theorem 1.6 (Strong Law of Large Numbers). *Let (X_i) be a sequence of iid random variables such that $\mathbb{E}\, X_i = \mu < \infty$. Then,*

$$n^{-1} \sum_{i=1}^{n} X_i \overset{a.s.}{\to} \mu.$$

1.2.2 Convergence in Probability

We now move to a weaker notion of convergence; *convergence in probability*, which we generally denote $X_n \overset{\mathbb{P}}{\to} X$.

Definition 1.7 (Convergence in probability). *A sequence (X_n) converges in probability to X if*

$$\forall \varepsilon > 0 : \mathbb{P}(|X_n - X| > \varepsilon) \overset{n\to\infty}{\to} 0.$$

Convergence in probability isn't expressed as naturally as an extension of the deterministic notion of convergence, but exploring why is illustrative.

Exercise #1: The Typewriter Sequence Suppose $X \sim \text{Uniform}(0,1)$ and define

$$Y_n = \mathbb{1}(X \in \left[\frac{n-2^k}{2^k}, \frac{n-2^k+1}{2^k}\right])$$

where, for each n, we choose $k = \lfloor \log_2(n) \rfloor$. For shorthand, we can write this as $Y_n = \mathbb{1}(X \in E_n)$ Show that $Y_n \xrightarrow{\mathbb{P}} 0$. As a bonus, show that $Y_n \overset{a.s.}{\not\to} 0$.

Solution: Let's first show that $Y_n \xrightarrow{\mathbb{P}} 0$. To do this, we want to show that for any $\varepsilon > 0$ there exists and $N \in \mathbb{N}$ such that $n \geqslant N \implies \mathbb{P}(|Y_n| > \varepsilon) \overset{n \to 0}{\to} 0$.

Fix $\varepsilon \in (0,1)$, then $\{|Y_n| > \varepsilon\} = \{Y_n = 1\}$. Then we have

$$\begin{aligned}
\mathbb{P}(|Y_n| > \varepsilon) &= \mathbb{P}(Y_n = 1) \\
&= \mathbb{P}(X \in E_n) \\
&= |E_n| \quad (X \text{ is uniform}) \\
&= 2^{-k}.
\end{aligned}$$

Because k is monotone non-decreasing to ∞ in n, we know that 2^{-k} goes to 0 (as an implicit function of n), and so $Y_n \xrightarrow{\mathbb{P}} 0$.

Now we'll show that $Y_n \overset{a.s.}{\not\to} 0$. Because $X \sim \text{Uniform}(0,1)$, when we view X as a function from $\Omega = (0,1)$ to $\mathbb{R}$ X becomes the identity function; $X(\omega) = \omega$ for $\omega \in (0,1)$. For any $\omega \in (0,1)$ and $N \in \mathbb{N}$, we know there exists an $n \geqslant N$ (in fact, there are infinitely many such n) such that $Y_n(\omega) = 1$. We can see this intuitively by observing that the events E_n look like a typewriter as a function of n (and implicitly k). The horizontal area spanned by the indicator for E_n scans across $[0,1]$ until it reaches the end, and then starts over, increasing k by 1 so it now operates at a finer resolution. There's a nice animated demonstration of this phenomenon here. On each full scan of $[0,1]$, there is guaranteed to be one value of n where $E_n = 1$ for any given ω.

Thus, $Y_n(\omega)$ cannot converge (in n) to 0 for any ω, so clearly Y_n does not converge almost surely. $\square$

This problem exhibits the sort of behavior that gives rise to a sequence which converges in probability but not almost surely. At any given n, there is a set of "bad" ω (in this case, ω such that $Y_n(\omega) = 1$). The set of bad ω changes as a function of n but does get less and less likely to be observed as n increases.

Just as we have a strong law of large numbers showing almost sure convergence, we have weak laws showing convergence in probability under two different sets of conditions.

Theorem 1.8 (Weak Law of Large Numbers)**.** *Suppose (X_i) is a sequence of iid r.v. with $\mathbb{E}\, X_1 < \infty$. Then,*

$$\bar{X}_n \xrightarrow{\mathbb{P}} \mathbb{E}\, X_1.$$

NOTE: You will often see the WLLN stated with an assumption that $\text{Var}(X_1) < \infty$ – this is a bit confusing (to me) because it's a stronger condition than the finite mean condition for the SLLN. This finite variance condition is not actually needed, but is often used to make it possible to prove using Chebyshev's. We'll revisit this idea later.

Theorem 1.9 (Weak Law of Large Numbers (for uncorrelated variables))**.** *Suppose (X_i) is a sequence of uncorrelated r.v. and define $\bar{X}_n = n^{-1}\sum_{i=1}^{n}$.*

If $\mathbb{E}\, \bar{X}_n \to \mu$ and $\limsup n^{-1} \sum_{i=1}^{n} Var(X_i) < \infty$, then

$$\bar{X}_n \xrightarrow{\mathbb{P}} \mu.$$

Note on Strong vs. Weak Laws: The language of "strong" and "weak" here refers to the type of convergence (almost sure vs. in probability), but there are many conditions under which one could get these convergence results. So, when we say "the" strong/weak law, we really mean "the most popular" strong/weak law.

1.2.3 Convergence in Distribution

Suppose a sequence $(X_n) \to X$. If it converges almost surely, this means that eventually $X_n(\omega) \approx X(\omega)$ for all ω in sets with non-zero probability. If it converges in probability, then eventually $X_n \approx X$ with high probability (over all possible sets of ω). An even weaker notion is to require only that the distribution of X_n gets close to the distribution of X:

Definition 1.10 (Convergence in distribution)**.** *For a r.v. Y we call F_Y its cdf. A sequence of r.v. (X_n) converges in distribution to X if*

$$F_{X_n}(x) \overset{n\to\infty}{\to} F_X(x)$$

at all $x \in \mathbb{R}$ where F_X is continuous.

Because the cdf is not well-defined for multivariate random vectors, we have the following more general definition:

Definition 1.11 (Convergence in distribution (more general))**.** *Let (X_n) be a sequence of r.v. with $X_n \in \mathcal{X}$. We say $X_n \overset{d}{\to} X$ if*

$$\mathbb{E}\, g(X_n) \to \mathbb{E}\, g(X)$$

for all bounded, continuous functions $g : \mathcal{X} \to \mathbb{R}$.

Yet another definition which will sometimes be analytically convenient is to use characteristic functions. Recall that the characteristic function of a r.v. X is given by $\phi_X(t) = \mathbb{E} \exp(itX)$.

Theorem 1.12 (Lévy's Continuity Theorem)**.** *For a sequence of r.v. (X_n) and r.v. X:*

$$X_n \overset{d}{\to} X \iff \phi_{X_n}(t) \to \phi_X(t) \text{ for all } t \in \mathbb{R}.$$

There are many ways to denote convergence in distribution, most of which involve either a d for "distribution" or $\mathcal{L}$ for "law". We'll use the notation $X_n \overset{d}{\to} X$.

The most famous result we have about convergence in distribution is the (Lindeberg-Levy) Central Limit Theorem. We'll give the more general multivariate version for k-dimensional random vectors, which easily recovers the one-dimensional version by setting $k = 1$.

Theorem 1.13 (Lindeberg-Levy Central Limit Theorem)**.** *Let (X_i) be a sequence of iid k-dimensional random vectors with $\Sigma = Cov(X_1)$ and $\Sigma_{ij} < \infty$ for all $i, j \in [k]$. Then,*

$$n^{-1/2} \sum_{i=1}^{n} (X_i - \mathbb{E}\, X_1) \overset{d}{\to} N(0, \Sigma).$$

Note that, with a bit of rearranging, we can state the LLNs as $\bar{X} - \mathbb{E}\, X_1 \to 0$ (either a.s. or in probablity) and the (one-dimensional) LL-CLT as $\sqrt{n}\left(\bar{X} - \mathbb{E}\, X_1\right) \overset{d}{\to} N(0, \mathrm{Var}(X_1))$.

Much like the case of two WLLNs above, there exist other CLTs which, generally speaking, relax an assumption on independence or identical distribution, at the cost of a stronger assumption elsewhere.

Theorem 1.14 (Lyapunov Central Limit Theorem)**.** *Let (X_i) be a sequence of independent r.v. with means $\mu_i = \mathbb{E}\, X_i$ and variances $\sigma_i^2 = Var(X_i) < \infty$. Define $s_n^2 = \sum_{i=1}^n \sigma_i^2$.*

If there exists some $\delta > 0$ such that

$$s_n^{-(2+\delta)} \sum_{i=1}^{n} \mathbb{E}\, |X_i - \mu_i|^{2+\delta} \to 0,$$

then,

$$s_n^{-1} \sum_{i=1}^{n} (X_i - \mu_i) \overset{d}{\to} N(0, 1).$$

1.2.4 Convergence in L^p

Definition 1.15 (L^p convergence)**.** *Let (X_n) be a sequence of r.v. with $X_n \in L^p$ for $p \geqslant 1$. Then we say that $X_n \overset{L^p}{\to} X$ (i.e. X_n converges to X in L^p) if $X \in L^p$ and*

$$\mathbb{E}\, |X_n - X|^p \to 0.$$

Note that $X_n \overset{L^p}{\to} X \implies \mathbb{E}\, |X_n|^p \to \mathbb{E}\, |X|^p$.

1.2.5 Useful facts about convergence

Fact 1.16 (Relationship between modes of convergence)**.**

$$X_n \xrightarrow{a.s.} X \implies X_n \xrightarrow{\mathbb{P}} X \implies X_n \xrightarrow{d} X$$
$$X_n \xrightarrow{L^p} \implies X_n \xrightarrow{\mathbb{P}} X \implies X_n \xrightarrow{d} X$$

It may be that we know about the convergence of a sequence X_n but want to know whether a function of X_n converges. For continuous functions, we have a very clean result.

Theorem 1.17 (Continuous Mapping Theorem)**.** *Let (X_i) be a sequence of r.v. with $X_i \in \mathcal{X}$ and let $g : \mathcal{X} \to \mathbb{R}$ be almost surely continuous (i.e. the set $D \subset \mathcal{X}$ where g is discontinuous is such that $\mathbb{P}(X \in D) = 0$). Then,*

$$X_n \xrightarrow{a.s.} X \implies g(X_n) \xrightarrow{a.s.} g(X_n)$$
$$X_n \xrightarrow{\mathbb{P}} X \implies g(X_n) \xrightarrow{\mathbb{P}} g(X_n)$$
$$X_n \xrightarrow{d} X \implies g(X_n) \xrightarrow{d} g(X_n)$$

Theorem 1.18 (Slutsky's Theorem)**.** *Let $(X_n), (Y_n)$ be sequences of r.v. such that $X_n \xrightarrow{d} X$ for some r.v. X and $Y_n \xrightarrow{\mathbb{P}} c$ for a constant c. Then,*

$$X_n + Y_n \xrightarrow{d} X + c$$
$$X_n Y_n \xrightarrow{d} Xc$$
$$\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{c} \text{ if } c \neq 0.$$

Theorem 1.19 (Delta Method)**.** *Let (X_n) be a sequence of r.v. such that*

$$\sqrt{n}(X_n - \theta) \xrightarrow{d} N(0, \sigma^2)$$

for some $\theta \in \mathbb{R}, \sigma^2 > 0$. Then for any g where $g(\theta \neq 0)$ and $g'(\theta)$ exists, we have

$$\sqrt{n}(g(X_n) - g(\theta)) \xrightarrow{d} N(0, g'(\theta)^2\sigma^2).$$

Exercise #2 Suppose (X_i) is an iid sequence with $X_i \in [1, \infty)$ and $\mu = \mathbb{E}\, X_i < \infty$. Define $T_n = \left(\prod_{i=1}^n X_i\right)^{1/n}$. Does T_n converge almost surely to a finite limit? If so, prove it. If not, prove that it does not.

Solution: Define $Y_i = \log(X_i)$ and $S_n = n^{-1}\sum_{i=1}^n Y_i$. Note that the Y_i are iid (because the X_i are iid). Moreover, we know $\mathbb{E}\,|Y_i| < \infty$ because $X_i > 1 \implies Y_i > 0$ and $Y_i < X_i \implies \mathbb{E}\, Y_i < \mathbb{E}\, X_i < \infty$. So, we can appeal to the strong law of large numbers to say that $S_n \xrightarrow{a.s.} \mathbb{E}\, Y_i$.

Now, by the continuous mapping theorem we know that $g(S_n) \xrightarrow{a.s.} g\left(\mathbb{E}\, Y_i\right)$ for any almost surely continuous g. Define $g(s) = \exp(s)$. We know the exponential function is continuous everywhere, so it must be almost surely continuous with respect to S_n. Therefore, we have $\exp\left(S_n\right) \xrightarrow{a.s.} \exp\left(\mathbb{E}\, Y_i\right)$.

Now, note that

$$\begin{aligned}
\exp(S_n) &= \exp\left(n^{-1}\sum_{i=1}^n Y_i\right) \\
&= \left(\prod_{i=1}^n \exp(Y_i)\right)^{1/n} \\
&= \left(\prod_{i=1}^n X_i\right)^{1/n} \\
&= T_n,
\end{aligned}$$

and so we have $T_n \xrightarrow{a.s.} \exp(\mathbb{E}\, Y_i)$. By Jensen's we know that $\exp(\mathbb{E}\, Y_i) \leqslant \mathbb{E} \exp(Y_i) = \mathbb{E}\, X_i < \infty$, and so T_n converges almost surely to a finite limit. □

Exercise #3: Proving WLLN with finite variance Prove the WLLN from Theorem 1.9, making the extra assumption that $\mathrm{Var}(X_1) < \infty$.
Hint: Start by showing the asymptotic behavior of $|\bar{X}_n - \mathbb{E}\, \bar{X}_n|$.

Solution: We'll start by showing that $(\bar{X}_n - \mathbb{E}\, \bar{X}_n) \xrightarrow{\mathbb{P}} 0$. Let $\varepsilon > 0$ be arbitrary, then

$$\begin{aligned}
\mathbb{P}(|\bar{X}_n - \mathbb{E}\, \bar{X}_n| > \varepsilon) &\leqslant \frac{\mathrm{Var}(\bar{X}_n)}{\varepsilon^2} \quad \text{(Chebyshev's)} \\
&= \frac{n^{-2} \sum_{i=1}^n \sum_{j=1}^n \mathrm{Cov}(X_i, X_j)}{\varepsilon^2} \\
&= \frac{n^{-1} \sum_{i=1}^n \mathrm{Var}(X_i)}{n\varepsilon^2} \quad (i \neq j \text{ terms are 0 bc of no correlation}) \\
&\xrightarrow{\mathbb{P}} 0 \quad (\text{numerator is finite, denominator} \to \infty)\,.
\end{aligned}$$

Now, we know that $\mathbb{E}\, \bar{X}_n \to \mu$ – note that this is deterministic convergence and thus, trivially, convergence in probability. So, by Slutsky's we have

$$\begin{aligned}
\bar{X}_n &= (\bar{X}_n - \mathbb{E}\, \bar{X}_n) + \mathbb{E}\, \bar{X}_n \\
&\xrightarrow{\mathbb{P}} 0 + \mu \\
&= \mu.
\end{aligned}$$

□

Exercise #4 Let (X_n) be a sequence of r.v. Prove that if $X_n \xrightarrow{d} c$ for some constant $c \in \mathbb{R}$, then $X_n \xrightarrow{\mathbb{P}} c$.

Solution: We start with a definition of convergence in distribution. Let F_{X_n} be the cdf of X_n and F_c be the cdf of the constant function $c(x) = c$. Then,

$$F_{X_n}(x) \to F_c(x)$$

for all x at which $F_c(x)$ is continuous. Note that

$$F_c(x) = \begin{cases} 0, & \text{if } x < c \\ 1, & \text{otw,} \end{cases}$$

so F_c is continuous at all $x \neq c$. In particular then, for any $\varepsilon > 0$ we know from $X_n \xrightarrow{d} c$ that

$$\begin{aligned}
F_{X_n}(c - \varepsilon/2) &\to 0 \\
F_{X_n}(c + \varepsilon/2) &\to 1.
\end{aligned}$$

We know that $\mathbb{P}(X_n \in [c - \varepsilon/2, c + \varepsilon/2]) = F_{X_n}(c + \varepsilon/2) - F_{X_n}(c - \varepsilon/2)$ and that $\lim(a_n + b_n) = \lim a_n + \lim b_n$, so we get

$$\begin{aligned}
\mathbb{P}\left(|X_n - c| > \varepsilon\right) &= 1 - \mathbb{P}\left(|X_n - c| \leqslant \varepsilon\right) \\
&\leqslant 1 - \mathbb{P}\left(X_n \in [c - \varepsilon/2, c + \varepsilon/2]\right) \quad \left(\{X_n \in [c - \varepsilon/2, c + \varepsilon/2]\} \subseteq \{|X_n - c| \leqslant \varepsilon\}\right) \\
&= 1 - F_{X_n}(c + \varepsilon/2) + F_{X_n}(c - \varepsilon/2) \\
&\to 1 - 1 + 0 \\
&= 0.
\end{aligned}$$

This is exactly the definition of $X_n \xrightarrow{\mathbb{P}} c$, so we're done. $\square$

Exercise #5: Proving WLLN without finite variance Prove the WLLN from Theorem 1.9 without assuming that $\text{Var}(X_1) < \infty$.
Hint: Use characteristic functions. You can take as given that, for X with $\mathbb{E}\, X = \mu < \infty$, we have

$$\begin{aligned}
\phi_{n^{-1}X}(t) &= \phi_X\left(n^{-1}t\right) \quad \text{(Hint \# 1)} \\
\phi_{X+Y}(t) &= \phi_X(t)\phi_Y(t) \quad \text{(Hint \# 2)} \\
\phi_X(t) &= 1 + it\mu + o(t) \quad \text{(Hint \# 3)} \\
\left(1 + i\mu t n^{-1} + o(tn^{-1})\right)^n &\overset{n\to\infty}{\to} \exp(it\mu) \quad \text{(Hint \# 4)}
\end{aligned}$$

Solution: Our first goal is to show that $\bar{X}_n \xrightarrow{d} \mu$, which we'll do by showing pointwise convergence of the characteristic function. Let t be arbitrary:

$$\begin{aligned}
\phi_{\bar{X}_n}(t) &= \phi_{n^{-1}\sum_{i=1}^n X_i}(t) \quad \text{(defn of } \bar{X}_n\text{)} \\
&= \prod_{i=1}^n \phi_{X_i}(tn^{-1}) \quad \text{(Hint \#1, \#2)} \\
&= \phi_{X_1}(tn^{-1})^n \quad (X_i \text{ are iid}) \\
&= \left(1 + i\mu t n^{-1} + o(tn^{-1})\right)^n \quad \text{(Hint \#3)} \\
&\overset{n\to\infty}{\to} \exp\left(it\mu\right) \quad \text{(Hint \#4)} \\
&= \phi_\mu(t).
\end{aligned}$$

So, we have pointwise convergence of the characteristic function, and thus $\bar{X}_n \xrightarrow{d} \mu$. μ is a constant, and so by the previous exercise we know that $\bar{X}_n \xrightarrow{\mathbb{P}} \mu$, which is our desired result. $\square$

2 Stochastic Processes (BST 230 Module 14)

2.1 Definition and Tools

Definition 2.1 (Stochastic Process)**.** *For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, measurable space (S, Σ), and index set T, a stochastic process is a collection of r.v.*

$$\{X_t : t \in T\}$$

where $X_t : \Omega \to S$.

An alternate viewpoint which is sometimes used is to think of the stochastic process in terms of each $\omega \in \Omega$ having a "sample path" $t \mapsto X(t, \omega)$, which shows how, for a fixed ω, our outcome $X_t(\omega)$ evolves over time.

Before we dive further into stochastic processes, it will be useful to developing a unifying language we can use to analyze many stochastic processes.

Definition 2.2 (Generating Function)**.** *The generating function of a sequence of real numbers $a = \{a_0, a_1, \ldots\}$ is*

$$G_a(s) = \sum_{n=0}^{\infty} a_n s^n.$$

At times we may want to take two sequences $(a_n), (b_n)$ and create a new sequence (c_n) where $c_n = \sum_{k=0}^n a_k b_{n-k}$. We call c_n the *convolution* of a_n, b_n and denote it as $c = a * b$.

Theorem 2.3 (Convolution Formula)**.** *For sequences $a = (a_0, a_1, \ldots)$ and $b = (b_0, b_1, \ldots)$ we have $G_{a*b}(s) = G_a(s)G_b(s)$.*

For a discrete random variable X taking values in a countable set $\mathcal{X} = (x_1, x_2, \ldots)$, one can imagine letting $a_n = \mathbb{P}(X = x_n)$, which gives rise to the idea of a *probability-generating function.*

Definition 2.4 (Probability-Generating Function)**.** *Let X be a random variable taking values in $\mathcal{X} = (x_1, x_2, \ldots)$. Then we call $G_X(s) = \sum_{n=1}^{\infty} \mathbb{P}(X = x_n)s^n$ the probability-generating function of X.*

Equivalently, one can think of $G_X(s) = \mathbb{E}(s^X)$ for $s \in [-1, 1]$.

Note the abuse of notation here; instead of subscripting the generating function with the set of real numbers, we subscript it with the random variable whose pmf is implied by this set (along with the space $\mathcal{X}$). In practice, we will often deal with random variables who take values either on $\mathbb{Z}$ or $\mathbb{N}$.

Probability-generating functions have a few useful properties. To simplify, let's assume that $\mathcal{X} = \mathbb{N}$ such that X takes only non-negative integral values.

- $G_X(s)$ is infinitely differentiable on $(-1, 1)$ and $\mathbb{P}(X = k) = \frac{G^{(k)}(0)}{k!}$
- $G_X(1) = 1$
- $\mathbb{E}\, X = G'_X(1)$
- $\mathrm{Var}(X) = G''_X(1) + G'_X(1)(1 - G'_X(1))$
- $\mathbb{E}\, \frac{X!}{(X-k)!} = G_X^{(k)}(1)$

It turns out that convolutions can help us understand sums of independent r.v.

Theorem 2.5 (PGF of sum of independent r.v.)**.** *Let X, Y be independent r.v. defined on $\mathbb{N}$ with pmfs p_X, p_Y. Then $G_{X+Y}(s) = G_{X*Y}(s) = G_X(s)G_Y(s)$.*

This fact can be generalized to the case where we have a sum of r.v. where the number of items being summed is itself a random variable.

Theorem 2.6 (Compounding Formula)**.** *Let (X_i) be a sequence of iid r.v. with a shared pgf $G_X(s)$ and $N \in \mathbb{N}$ be a random variable with pgf $G_N(s)$. Then $Z_N = \sum_{i=1}^{N} X_i$ has the pgf*

$$G_{Z_N}(s) = G_N\left(G_X(s)\right).$$

Exercise #6: Wald's Identity Using the random variables defined in the theorem above, show that

$$\mathbb{E}\, Z_N = \mathbb{E}[N]\, \mathbb{E}[X_1].$$

NOTE: this is a very useful fact to know

Solution: This follows from the rules we've defined above and an application of the chain rule.

$$\begin{aligned}
\mathbb{E}\, Z_N &= G'_{Z_N}(1) \\
&= G'_N(G_{X_1}(1))G'_{X_1}(1) \quad \text{(chain rule: } (f(g(x)))' = f'(g(x))g'(x)\text{)} \\
&= G'_N(1)G'_{X_1}(1) \quad (G_{X_1}(1) = 1 \text{ by construction of pgf}) \\
&= \mathbb{E}[N]\, \mathbb{E}[X_1]
\end{aligned}$$

□

2.2 Branching Processes (Galton-Watson)

Branching processes model the number of individuals in a population over a number of generations. We assume that individuals survive for only one generation.

- Z_n: number of individuals at generation n (assume $Z_0 = 1$)
- $X_{n,i}$: number of offspring had by individual i in generation n
- $(X_{n,i})$ are iid (over both n and i) with pgf $G_X(s)$

Note that $Z_{n+1} = \sum_{i=1}^{Z_n} X_{n,i}$, so by the compounding formula and iteration, $G_{Z_{n+1}} = G_{Z_n}(G_X(s)) = \ldots$. Iterating further tells us that G_{Z_n} is given by the n-fold composition of G_x.

Exercise #7 Show that if each individual is expected to have μ offspring, then $\mathbb{E}\, Z_n = \mu^n$.

Solution:

$$\begin{aligned}\mathbb{E}\, Z_n &= \mathbb{E}\left(\sum_{i=1}^{Z_{n-1}} X_{(n-1),i}\right) \\ &= \mathbb{E}[Z_{n-1}]\mu \quad \text{(Wald's identity)}\end{aligned}$$

Then, by an inductive argument carrying this out n times, we get $\mathbb{E}\, Z_n = \mu^n$. $\square$

We may also be interested in the probability of "ultimate extinction"; that is, $\lim_{n\to\infty} \mathbb{P}(Z_n = 0)$. The probability of ultimate extinction is 1 if $\mu \leqslant 1$ and is < 1 if $\mu > 1$.[1]

2.3 Poisson Processes

Definition 2.7 (Poisson Process)**.** *Let (X_i) be a sequence of iid r.v. with $X_i \sim Exp(\lambda)$ and define $T_n = \sum_{i=1}^n X_i$.*

Define $Y_t = \sum_{n=1}^\infty \mathbb{1}(T_n \leqslant t)$ for $t \geqslant 0$. Then we call the continuous-time stochastic process (Y_t) a Poisson process with arrival times $(T_1, T_2, \ldots)$.

We call this a Poisson process because $Y_t \sim \text{Pois}(\lambda t)$. If we define $t_0 < t_1 < \ldots < t_k$ and $W_i = Y_{t_i} - Y_{t_{i-1}}$, then $W_i \perp W_j$ for $i \neq j$ and $W_i \sim \text{Pois}(\lambda(t_i - t_{i-1}))$. That is, W_i is the number of arrivals in the interval $(t_{i-1}, t_i]$ and this number is Poisson, scaled by the width of this interval (i.e. the amount of time being considered) as well as the baseline rate λ. Moreover, for two disjoint time intervals, the number of arrivals in the second interval is independent of the number of arrivals in the first.

2.4 Martingales

Martingales are a class of stochastic processes which are both reasonably simple to define and encompass many other types of stochastic processes. This provides potentially huge benefits to us when confronted with a new stochastic process; if we can represent it as a martingale, we immediately get a large number of powerful results for analyzing it.

Martingales come in both discrete-time and continuous-time varieties, but we'll focus on the discrete-time case.

Definition 2.8 (Martingale)**.** *A discrete time stochastic process (Y_n) is a martingale with respect to another process (X_n) if*

$$\sup_n \mathbb{E}\,|Y_n| < \infty$$
$$\mathbb{E}\, Y_{n+1} | X_{1:n} = Y_n \text{ for all } n \in \mathbb{N}.$$

These simple conditions are already enough to guarantee almost sure convergence of (Y_n) to some r.v. Y.

Theorem 2.9 (Martingale Convergence Theorem)**.** *If (Y_n) is a martingale, there exists a r.v. Y such that $Y_n \xrightarrow{a.s.} Y$.*

If $\limsup_n \mathbb{E}\, Y_n^2 < \infty$, the we also know that $Y_n \xrightarrow{L^2} Y$.

[1] Technically, this assumes some randomness in the number of offspring. If each individual always has exactly one offspring, then the probability of extinction is 0 even though $\mu = 1$.

Exercise #8 Let (X_n) be a sequence of r.v. where $X_0 = \frac{1}{2}$ and $X_n \sim \text{Unif}(0, 2X_{n-1})$ for $n \geqslant 1$. Assume the randomness in this uniform is independent; in other words, $Y_n = \frac{X_n}{2X_{n-1}} \overset{iid}{\sim} \text{Unif}(0,1)$.

(a) Show that $X_n \overset{a.s.}{\to} 0$.

Hint: Show that (X_n) is a martingale where each element can be written as $X_n = \prod_{i=1}^n U_i$ where $U_i \overset{iid}{\sim} \text{Unif}(0,1)$, and show that it converges almost surely to some r.v. Then use non-martingale techniques to show that $X_n \overset{a.s.}{\to} 0$.

(b) Does $X_n \overset{L^1}{\to} 0$? If so, prove it. If not, prove that it does not.

NOTE: This is not the greatest example because the martingale step isn't really necessary, but it is a generally useful tool worth practicing.

Solution:

Part (a1): form of X_n By the scaling properties of the uniform distribution, it is immediate that $X_n = 2X_{n-1}U_n$ where $U_n \sim \text{Unif}(0,1)$. Then, by induction, we see that $X_n = 2^n X_0 \prod_{i=1}^n U_n = 2^{n-1}\prod_{i=1}^n U_i$ where the U_i are independent.

Part (a2): Show (X_n) is martingale Now let's show that (X_n) is a martingale. Because X_n is a product of elements $\in (0,1)$, $X_n \in (0,1)$ and so $\sup_n \mathbb{E}|X_n| \leqslant 1 < \infty$. Then, note that we can write

$$
\begin{aligned}
\mathbb{E}\, X_{n+1}|X_{1:n} &= \mathbb{E}\, X_{n+1}|X_n \quad \text{(}(X_n)\text{ is a Markov chain)} \\
&= \mathbb{E}\left[\text{Unif}(0, 2X_n)|X_n\right] \\
&= X_n.
\end{aligned}
$$

So, (X_n) is a martingale. Thus, by the martingale convergence theorem we know that $X_n \overset{a.s.}{\to} X$ for some r.v. X. It now just remains to show that $X = 0$.

Part (a3): Show $X_n \overset{a.s.}{\to} 0$ Let's consider $\log X_n = \log\left(2^{n-1}\prod_{i=1}^n U_i\right) = (n-1)\log 2 + \sum_{i=1}^n \log(U_i)$. By the strong law of large numbers, we know that $n^{-1}\sum_{i=1}^n \log U_i \overset{a.s.}{\to} \mathbb{E}\log(U_i) = -1$. We calculate this -1 by noting that $\mathbb{E}\log(U_i) = \int_0^t \log(u)du = -1$. Thus, we have

$$
\begin{aligned}
\log X_n &= (n-1)\log 2 + \sum_{i=1}^n \log(U_i) \\
&= n\left(\underbrace{\frac{n-1}{n}\log 2}_{\overset{a.s.}{\to}\log 2} + \underbrace{\frac{1}{n}\sum_{i=1}^n \log(U_i)}_{\overset{a.s.}{\to}-1}\right) \\
&\overset{a.s.}{\to} -\infty \quad \left(\text{because } n\cdot c \overset{n\to\infty}{\to} -\infty \text{ for } c<0 \text{ and } \log 2 - 1 < 0\right)
\end{aligned}
$$

The $\exp(\cdot)$ function is continuous, so the continuous mapping theorem tells us that $X_n = \exp(\log X_n) \overset{a.s.}{\to} \exp(-\infty) = 0$.

Part (b): Does $X_n \overset{L^1}{\to} 0$? No, we can see this using the law of total expectation and martingale property inductively:

$$
\begin{aligned}
\mathbb{E}\, X_n &= \mathbb{E}\left[\mathbb{E}[X_n|X_{1:(n-1)}]\right] \quad \text{(law of total expectation)} \\
&= \mathbb{E}\left[X_{n-1}\right] \quad \text{(martingale property)} \\
&= \mathbb{E}\left[\mathbb{E}[X_{n-1}|X_{1:(n-2)}]\right] \\
&= \mathbb{E}\left[X_{n-2}\right]
\end{aligned}
$$

$$
\begin{aligned}
&\vdots \\
&= \mathbb{E}\, X_0 \\
&= \frac{1}{2}.
\end{aligned}
$$

So $\mathbb{E}\, X_n = \frac{1}{2}$ for all X_n. Recall that $X_n \xrightarrow{L^p} X \implies \mathbb{E}\,|X_n|^p \to \mathbb{E}\,|X|^p$. So, $X_n \xrightarrow{L^1} 0 \implies \mathbb{E}\,|X_n| \to 0$. However, we see that $\mathbb{E}\, X_n = \mathbb{E}\,|X_n| = \frac{1}{2}$, so we have a contradition and know that X_n cannot converge in L^1 to 0. $\square$

3 Extra Exercises

Exercise #9 Let (X_n) be an iid sequence with $\mathbb{E}\, X_1 = 0$ and $\mathrm{Var}(X_1) = \infty$.
Prove that

$$\mathbb{P}\left(\limsup_n \{|X_n| \geqslant \sqrt{n}\}\right) = 1.$$

Hint: For any non-negative r.v. Y, $\mathbb{E}\, Y = \int_{t=0}^{\infty} \mathbb{P}(Y \geqslant t)dt$ (This is called the "layer cake representation" of a non-negative discrete random variable).
Note: This is the set-theoretic notion of "limsup" – i.e. $\omega \in \limsup_n E_n \iff \omega \in E_n$ for infinitely many n. So, the thing you're proving here is that, almost surely, $X_n \geqslant \sqrt{n}$ for infinitely many n.

Solution: Because the X_n are independent, the events $\{|X_n| \geqslant \sqrt{n}\}$ are independent, and we can appeal to the second Borel-Cantelli lemma. It will suffice to show that $\sum_{n=1}^{\infty} \mathbb{P}(|X_n| \geqslant \sqrt{n}) = \infty$.

Note that

$$
\begin{aligned}
\sum_{n=1}^{\infty} \mathbb{P}(|X_n| \geqslant \sqrt{n}) &= \sum_{n=1}^{\infty} \mathbb{P}\left(X_n^2 \geqslant n\right) \\
&= \sum_{n=1}^{\infty} \mathbb{P}\left(X_1^2 \geqslant n\right) \quad (X_i \text{ are iid})
\end{aligned}
$$

Assume toward a contradiction that $\sum_{n=1}^{\infty} \mathbb{P}\left(X_1^2 \geqslant n\right) < \infty$. Then, by the hint we have

$$
\begin{aligned}
\mathbb{E}\, X_n^2 &= \int_{t=0}^{\infty} \mathbb{P}(X_1^2 > t)dt \\
&\leqslant \sum_{n=0}^{\infty} \mathbb{P}\left(X_1^2 > n\right) \quad \text{(see below)} \\
&\leqslant \sum_{n=0}^{\infty} \mathbb{P}\left(X_1^2 \geqslant n\right) \\
&\leqslant 1 + \sum_{n=1}^{\infty} \mathbb{P}\left(X_1^2 \geqslant n\right) \\
&< \infty.
\end{aligned}
$$

As an aside, note that the second line follows because

$$\int_{t=0}^{\infty} \mathbb{P}(X_1^2 > t)dt \leqslant \int_{t=0}^{\infty} \mathbb{P}(X_1^2 > \lfloor t \rfloor)dt = \sum_{n=0}^{\infty} \mathbb{P}(X_1^2 > n).$$

However, we know that $\mathbb{E}\, X_1^2 = \mathrm{Var}(X_1) = \infty$, so we have a contradiction. Therefore, it must be that $\sum_{n=1}^{\infty} \mathbb{P}\left(X_1^2 \geqslant n\right) = \infty$.

So, by the second Borel-Cantelli lemma, we know that

$$\mathbb{P}\left(\limsup_n \{|X_n| \geqslant \sqrt{n}\}\right) = 1.$$

□