Review Session 5 – Stochastic Convergence

References/suggested reading

(i) Casella & Berger, section 5.5.

1 Introduction

In the previous review session, we introduced the notion of a random sample of size n, and the beginnings of *estimation* where we try to deduce probable information about the underlying distribution based upon a sample. Intuitively, as n becomes larger, we should obtain more information about the distribution. We'd like to understand the extent to which this is true and so we will need to develop a notion of convergence for random variables. Many of the rules you learned about for the classical convergence/limit of real sequences will hold here.

The limiting behavior of an estimator or procedure is a good approximation for the behavior when the sample size n is large, and can often even inform us of the general finite-sample case (or at least, the two are often conflated in practice). Thus, *asymptotic analysis* is a central part of statistical theory.

2 Convergence in Probability

We start with a notion of convergence based on the intuition that if X_n "converges" to X, then the mass of large deviations "$|X_n - X| > \epsilon$" should become smaller as n goes to 0.

Definition 2.1 (convergence in probability)**.** A sequence of random variables $X_1, X_2, \ldots$ *converges in probability* to a random variable X if, for every $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \epsilon) = 0,$$

or equivalently $\lim_n \mathbb{P}(|X_n - X| < \epsilon) = 1$. We will abbreviate this as $X_n \xrightarrow{\mathsf{P}} X$.

Theorem 2.2 (weak law of large numbers)

Let $X_1, X_2, \ldots$ be i.i.d. random variables with $\mathbb{E}[X_i] = \mu$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Define $\overline{X}_n := (1/n)\sum_{i=1}^n X_i$. Then, $X_n \xrightarrow{\mathsf{P}} \mu$.

Proof. We have for every $\epsilon > 0$, by Markov's ienquality:

$$\mathbb{P}(|\overline{X}_n - \mu| \geq \epsilon) = \mathbb{P}((\overline{X}_n - \mu)^2 \geq \epsilon^2) \leq \frac{\mathrm{Var}(\overline{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n \cdot \epsilon^2} \overset{n\to\infty}{\longrightarrow} 0.$$

■

Example 2.3 (consistency of S^2)
Suppose we have a sequence $X_1, X_2, \ldots$ of iid random variables with $\mathbb{E}(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$. If we define

$$S_n^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

can we prove a WLLN for S_n^2? Using Chebyshev's Inequality, we have

$$\mathbb{P}(|S_n^2 - \sigma^2| \geq \epsilon) \leq \frac{\mathbb{E}(S_n^2 - \sigma^2)^2}{\epsilon^2} = \frac{\mathrm{Var}\, S_n^2}{\epsilon^2}$$

and thus, a sufficient condition that S_n^2 converges in probability to σ^2 is that $\mathrm{Var}\, S_n^2 \to 0$ as $n \to \infty$.

Theorem 2.4 (continuous mapping theorem)
Suppose that $X_n \xrightarrow{\mathsf{P}} X$ and that h is a continuous function. Then, $h(X_n) \xrightarrow{\mathsf{P}} h(X)$.

Example 2.5 (consistency of $S = \sqrt{S^2}$)
By the continuous mapping theorem, we have the sample standard deviation $S_n = \sqrt{S_n^2} = h(S_n^2)$ is a consistent estimator of σ.

3 Almost Sure Convergence

Almost sure convergence (also called "almost everywhere convergence", or "convergence with probability 1", or "strong convergence") is another notion of convergence that is stronger than the previously mentioned convergence in probability. This type of convergence is similar to pointwise convergence of a sequence of functions. Recall that we defined a random variable X as really a function $X : \mathcal{S} \to \mathbb{R}$ from a sample space $\mathcal{S}$. Then, we can think about convergence at each element of the sample space: $\forall s \in \mathcal{S} : X_n(s) \to X(s)$. There is one caveat here, contained in the word "almost". We don't care about the value of random variables on sets which occur with probability 0. Thus, we should only be concerned with convergence on sets with probability 1.

Definition 3.1 (almost sure convergence)**.** A sequence of random variables $X_1, X_2, \ldots$ *converges almost surely*, or $X_n \xrightarrow{\text{a.s.}} X$ if, for every $\epsilon > 0$:

$$\mathbb{P}\left(\lim_n |X_n - X| < \epsilon\right) = 1.$$

This is equivalent to $\mathbb{P}\left(s \in \mathcal{S} : \lim_n X_n(s) = X(s)\right) = 1$.

Example 3.2 (almost sure but not sure convergence)
Let the sample space $\mathcal{S}$ be the closed interval $[0, 1]$ with the uniform probability distribution. Define random variables $X_n(s) = s + s^n$ and $X(s) = s$. For every $s \in [0, 1)$, $s^n \to 0$ as $n \to \infty$ and $X_n(s) \to s = X(s)$. However, $X_n(1) = 2$ for every n so $X_n(1)$ does not converge to $1 = X(1)$. But since the convergence occurs on the set $[0, 1)$ and $\mathbb{P}([0, 1)) = 1$, X_n converges almost surely to X.

Remark 3.3. Almost sure convergence implies convergence in probability.

Example 3.4 (convergence in probability, but not almost surely)

Again, let the sample space $\mathcal{S}$ be the closed interval $[0,1]$ with the uniform probability distribution. Now, define the sequence of real numbers $\{a_n\}$ by $a_n := \frac{n-2^k}{2^k}$ where $k \in \mathbb{N} \cup \{0\}$ is the largest integer such that $2^k \le n$. Then, define the sequence of random variables $X_1, X_2, \dots$ as $X_n(s) := \mathbf{1}\{s \in [a_n, a_{n+1})\}$. Then, the sequence $\{X_n\}$ graphically looks like a "travelling block which gets thinner and thinner". $X_n \xrightarrow{\text{P}} 0$ since the set of s's for which $X_n(s) \ge \epsilon$ gets thinner and thinner as $n \to \infty$. On the other hand, there is no value of $s \in \mathcal{S}$ for which $X_n(s) \to 0$ since for any $\epsilon \in (0,1)$, there is an arbitrarily large $n \in \mathbb{N}$ for which $X_n(s) > \epsilon$ (i.e., the blocks stay the same height).

Theorem 3.5 (continuous mapping theorem)

Suppose that $X_n \xrightarrow{\text{a.s.}} X$ and that h is a continuous function. Then, $h(X_n) \xrightarrow{\text{a.s.}} h(X)$.

Proof. This follows from the definition of a continuous function. ∎

Theorem 3.6 (strong law of large numbers)

Let $X_1, X_2, \dots$ be i.i.d. random variables with $\mathbb{E}[X_i] = \mu < \infty$, and define $\overline{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$. Then, $\overline{X}_n \xrightarrow{\text{a.s.}} \mu$.

Proof. We give an idea of the proof under the stronger assumption that X_i has finite fourth moment: $\mathbb{E}[X_i^4] < \infty$. WLOG, suppose $\mu = 0$ since we can just subtract μ from each X_i and appeal to the continuous mapping theorem. Let $S_n := \sum_{i=1}^n X_i$. Then, expanding the fourth power of a sum gives us:

$$\mathbb{E}[S_n^4] = n \cdot \mathbb{E}[X_1^4] + 3(n^2 - n) \cdot (\mathbb{E}[X_1^2])^2,$$

where we used the fact that $\mu = 0$. In other words $\mathbb{E}[S_n^4] \le Cn^2$ for some $C > 0$. Thus, by Chebyshev's inequality

$$\mathbb{P}(|S_n| > n\epsilon) = \mathbb{P}(|S_n^4| \ge n^4\epsilon^4) \le \frac{\mathbb{E}[|S_n^4|]}{n^4\epsilon^4} \le \frac{C}{n^2\epsilon^4}.$$

In other words, we've shown that the probability that the average deviates from 0 by more than ϵ, or $|S_n/n| > \epsilon$, is small and scales like n^{-2}. Then, we have

$$\sum_{n=1}^{\infty} \mathbb{P}(|S_n/n| > \epsilon) < \infty.$$

The proof is finished using a lemma called the Borel-Cantelli lemma which states that if the above holds, then

$$\mathbb{P}\left(\forall n \in \mathbb{N} : \exists k \ge n : |S_k/k| > \epsilon\right) = 0.$$

The event above should be read as: for every positive integer $n \in \mathbb{N}$, there is a $k \ge n$ for which the deviation is large $|S_k/k| > \epsilon$. This is exactly the negation of $S_k/k \to 0$. Thus, $\mathbb{P}\left(\lim_n |\overline{X}_n| > \epsilon\right) = 0$ and $\overline{X}_n \xrightarrow{\text{a.s.}} 0$. ∎

> **Remark 3.7.** WLLN in fact holds without the finite fourth moment assumption (even though our short proof used this assumption).

4 Convergence in Distribution

We have already encountered the idea of convergence in distribution in review session 2, where we defined convergence of a sequence of random variables by the pointwise convergence of their cdf's. This is also sometimes called *weak convergence* since it is implied by the other two modes of convergence.

Definition 4.1 (convergence in distribution)**.** A sequence of random variables $X_1, X_2, \dots$ *converges in distribution* to a random variable X if the cdf's converge, or

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x),$$

at all points x where $F_X(x)$ is continuous. We denote this by $X_n \xrightarrow{\text{d}} X$.

Example 4.2

Why might we only care about the points of continuity here? It turns out this is essential. Let's consider an example where X_n is uniform on the interval $(0, 1/n)$. If we look at the graph of the pdf of X_n, it seems like the mass of X_n gradually moves towards 0. So, X_n should converge to the random variable $X = 0$ (by which we mean $X = 0$ with probability 1). However, the cdf $F_n(x)$ of X_n vanishes at $x = 0$. So, we see $F_n(0) \overset{n\to\infty}{\to} 0$, but $F_X(0)$, the cdf of X evaluated at 0, is $1 \neq 0$. Thus, the convergence of cdf's fails at the point $x = 0$ where $F_X(\cdot)$ is discontinuous. However, we still have $X_n \xrightarrow{\text{d}} X$.

Example 4.3 (maximum of uniforms)

If $X_1, X_2, \ldots$ are iid uniform$(0, 1)$, we have

$$\mathbb{P}(|X_{(n)} - 1| \geq \epsilon) = \mathbb{P}(X_{(n)} \geq 1 + \epsilon) + \mathbb{P}(X_{(n)} \leq 1 - \epsilon) = \mathbb{P}(X_{(n)} \leq 1 - \epsilon)$$

which is

$$\mathbb{P}(X_{(n)} \leq 1 - \epsilon) = \mathbb{P}(X_i \leq 1 - \epsilon, i = 1, \ldots, n) = (1 - \epsilon)^n$$

which goes to 0 meaning $X_{(n)} \to 1$ in probability. However, if we take $\epsilon = t/n$, we have

$$\mathbb{P}(X_{(n)} \leq 1 - t/n) = (1 - t/n)^n \to e^{-t}$$

giving

$$\mathbb{P}(n(1 - X_{(n)}) \leq t) \to 1 - e^{-t}$$

Thus, the random variable $n(1 - X_{(n)})$ converges in distribution to an $\exp(1)$ random variable.

Theorem 4.4 (relation to other forms of convergence)

$X_n \xrightarrow{\text{P}} X \implies X_n \xrightarrow{\text{d}} X$.

Theorem 4.5

The sequence of random variables $\{X_n\}$ converges in probability to a constant c iff the sequence also converges in distribution to c. In order words: $X_n \xrightarrow{\text{d}} c \iff X_n \xrightarrow{\text{P}} c$.

Recall from review session 2 that convergence in distribution is characterized by convergence of the mgf's, if they exist.

Theorem 4.6 (Lévy continuity theorem for mgf's)

Suppose $\{X_i\}_{i=1}^{\infty}$ is a sequence of random variables, each with mgf $M_{X_i}(t)$. Furthermore, suppose that $\lim_{i\to\infty} M_{X_i}(t) = M(t)$ for all t in a neighborhood of 0 and $M(t)$ is an mgf. Then there is a unique cdf F_X whose moments are determined by $M_X(t)$ and, for all x where $F_X(x)$ is continuous, we have

$$\lim_{i\to\infty} F_{X_i}(x) = F_X(x)$$

That is, convergence for $|t| < h$, of mgfs to an mgf implies convergence in distribution.

Theorem 4.7 (central limit theorem)

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables. Let $\mathbb{E}[X_i] = \mu$ and $\mathrm{Var}(X_i) := \sigma^2 < \infty$. Then,

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{\text{d}} \mathcal{N}(0, 1).$$

Proof. We'll give an outline of the proof assuming X_i has an mgf. Of course, the mgf may not exist in general – but it can be replaced in this argument with the more general characteristic function, which always exists.

Let's start by assuming WLOG that X_i has mean 0 and variance 1. This is fine since $\frac{\overline{X}_n - \mu}{\sigma} = \frac{1}{n}\sum_{i=1}^n \frac{X_i - \mu}{\sigma}$, and thus we can standardize each X_i. Thus, it suffices to show $\sqrt{n}\cdot\overline{X}_n \xrightarrow{\text{d}} \mathcal{N}(0,1)$. We'll use Lévy's continuity theorem to achieve this. We have the mgf of $\sqrt{n}\cdot\overline{X}_n$ is

$$M_{\sqrt{n}\cdot\overline{X}_n}(t) = M_{\sum_{i=1}^n X_i}\left(\frac{t}{\sqrt{n}}\right) = \left(M_X\left(\frac{t}{\sqrt{n}}\right)\right)^n.$$

We now expand $M_X(t/\sqrt{n})$ in a Taylor series around 0:

$$M_X\left(\frac{t}{\sqrt{n}}\right) = 1 + \frac{(t/\sqrt{n})^2}{2!} + o(t^2/n).$$

Thus, we have

$$\left(M_X\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 + \frac{1}{n}\left(\frac{t^2}{2} + n\cdot o(t/n^2)\right)\right)^n \overset{n\to\infty}{\longrightarrow} \lim_n \left(1 + \frac{1}{n}\left(\frac{t^2}{2}\right)\right)^n = e^{t^2/2}.$$

In the above, we used the fact that the second-order Taylor expansion remainder, denoted by $o(t^2/n)$, goes to 0 faster than $(1/\sqrt{n})^2 = 1/n$. Thus, $n\cdot o(t/\sqrt{n}) \to 0$. ∎

Theorem 4.8 (portmanteau theorem)

$X_n \xrightarrow{\text{d}} X$ iff $\mathbb{E}[f(X_n)] \overset{n\to\infty}{\longrightarrow} \mathbb{E}[f(X)]$ for every f a bounded and continuous function.

Proof. We'll give a sketch of the proof. The reverse direction would follow from letting $f(x) := \mathbf{1}\{x \le t\}$ whence $\mathbb{E}[\mathbf{1}\{X \le t\}] = F_X(t)$, the cdf of X. Thus, the statement $\mathbb{E}[f(X_n)] \overset{n\to\infty}{\longrightarrow} \mathbb{E}[f(X)]$ gives us pointwise convergence of cdf's. There's only one problem with this argument! $f(x)$ is bounded, but not continuous. Thus, we have to take a continuous approximation $\tilde{f}(x)$ of $f(x)$ (imagine "smoothing out" an indicator function to make it continuous), and argue that $\mathbb{E}[\tilde{f}(X)] \to \mathbb{E}[f(X)]$ as the approximation $\tilde{f}$ converges to f.

The forward direction is a similar argument, but in reverse. If $X_n \xrightarrow{\text{d}} X$, then we know the desired convergence holds for functions f of the form $f(x) = \mathbf{1}\{x \le t\}$. By linearity of expectation and convergence, we can further show it holds for step functions or functions f of the form $f(x) = \sum_{i=1}^n a_i \cdot \mathbf{1}\{x \in (b_i, c_i)\}$. Then, it remains to show these step functions can approximate a general bounded and continuous function, in the sense that their integrals converge.

The details here are omitted, but this is a very standard type of approximation argument in probability theory, where if we want to prove something for a broad class of functions, we first focus on a simpler subclass of functions (i.e., indicator functions over intervals) where it is easier to prove and then move to the general case by taking another limit. ∎

Theorem 4.9 (continuous mapping theorem)

If $X_n \xrightarrow{\text{d}} X$, then $h(X_n) \xrightarrow{\text{d}} h(X)$ when h is a continuous function.

Proof. This follows from portmanteau theorem. We have that if $X_n \xrightarrow{\text{d}} X$, then $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ for every bounded and continuous f. In particular, for any bounded and continuous function f, $f \circ h$ is also a bounded and continuous function. Thus, $\mathbb{E}[f(h(X_n))] \to \mathbb{E}[f(h(X))]$ meaning $h(X_n) \xrightarrow{\text{d}} h(X)$ by portmantau theorem. ∎

Theorem 4.10 (Slutsky's theorem)

If $X_n \xrightarrow{\mathsf{d}} X$ and $Y_n \xrightarrow{\mathsf{P}} c$, a constant, (recall this is equivalent to saying $Y_n \xrightarrow{\mathsf{d}} c$) then

1. $Y_n X_n \xrightarrow{\mathsf{d}} c \cdot X$.
2. $X_n + Y_n \xrightarrow{\mathsf{d}} X + c$.

Proof. This follows from the continuous mapping theorem. ∎

Example 4.11 (normal approximation with estimated variance)

Suppose

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{\mathsf{d}} \mathcal{N}(0,1)$$

but that the value of σ is unknown. We have seen in a previous example that if $\lim_{n\to\infty} \operatorname{Var} S_n^2 = 0$, then $S_n^2 \to \sigma^2$ in probability. We can show this implies $\sigma / S_n \to 1$ in probability. Hence, by Slutsky,

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n} = \frac{\sigma}{S_n} \cdot \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{\mathsf{d}} \mathcal{N}(0,1)$$

5 Generalizations of CLT

Lyapunov's CLT is a more general version of the central limit theorem which holds for a sequence of random variables X_i which are independent, but not necessarily identically distributed. The cost of allowing for more general sequences of random variables $\{X_n\}$, however, is that we need some control on the average size of the deviations $X_n - \mathbb{E}[X_n]$ which can now vary greatly as n changes. One way of doing this is by establishing a bound on the higher moments of the deviation $|X_n - \mathbb{E}[X_n]|^{2+\delta}$.

Theorem 5.1 (Lyapunov CLT)

Suppose $X_1, X_2, \ldots$ is a sequence of independent random variables, each with finite expected value μ_i and variance σ_i^2. Define

$$s_n^2 := \sum_{i=1}^{n} \sigma_i^2.$$

If for some $\delta > 0$, *Lyapunov's condition*

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}[|X_i - \mu_i|^{2+\delta}] = 0,$$

is satisfied, then

$$\frac{1}{s_n} \sum_{i=1}^{n} (X_i - \mu_i) \xrightarrow{\mathsf{d}} \mathcal{N}(0,1).$$

Remark 5.2. In practice, it is often easiest to check Lyapunov's condition for $\delta = 1$ (or to compute the third moments of X_i).

Lyapunov's CLT can also be stated for *triangular arrays*. A triangular array is an array of random variables of the form

$$\begin{array}{ccccc} X_{1,1} & & & & \\ X_{2,1} & X_{2,2} & & & \\ X_{3,1} & X_{3,2} & X_{3,3} & & \\ \vdots & \vdots & \vdots & \ddots & \\ X_{n,1} & X_{n,2} & X_{n,3} & \cdots & X_{n,n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{array},$$

where each row $\{X_{n,1}, \ldots, X_{n,n}\}$ is an i.i.d. collection of random variables. The key generality here is that the distirbution is allowed to change from row to row. This is a situation which can arise, for example, when looking at a sequence of random matrices $\mathbf{X}_n \in \mathbb{R}^{n \times n}$ which grows in dimension with n.

Theorem 5.3 (Lyapunov CLT for triangular arrays)

Suppose $\{X_{n,k}\}_{n,k}$ is a triangular array where for each n, $X_{n,1}, X_{n,2}, \ldots$ are independent with $\mathbb{E}[X_{n,k}] = 0$ for $k = 1, 2, \ldots, n$. If $s_n^2 := \sum_{k=1}^n \sigma_{n,k}^2$, where $\mathrm{Var}(X_{n,k}) = \sigma_{n,k}^2$, and for some $\delta > 0$:

$$\lim_{n \to \infty} \sum_{k=1}^n \frac{1}{s_n^{2+\delta}} \mathbb{E}[|X_{n,k}|^{2+\delta}] = 0,$$

then

$$\frac{1}{s_n} \sum_{k=1}^n X_{n,k} \xrightarrow{\mathsf{d}} \mathcal{N}(0, 1).$$

Theorem 5.4 (multivariate CLT)

Let $\{\mathbf{X}_n\}$ be a sequence i.i.d. random vectors in $\mathbb{R}^p$ with mean μ and covariance Σ. Then,

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \mathbf{X}_i - \mu\right) \xrightarrow{\mathsf{d}} \mathcal{N}(\mathbf{0}_p, \Sigma).$$

Idea. By the univariate CLT, for any $\mathbf{a} \in \mathbb{R}^p$, we have

$$\mathbf{a}^T\left(\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \mathbf{X}_i - \mu\right)\right) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \mathbf{a}^T\mathbf{X}_i - \mathbf{a}^T\mu\right) \xrightarrow{\mathsf{d}} \mathcal{N}(0, \mathbf{a}^T\Sigma\mathbf{a}).$$

Thus, we see that any projection of the limiting distribution of $\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \mathbf{X}_i - \mu\right)$ is Gaussian. This means, by the Cramér-Wold device (see review session 4), the limiting distribution must also be Gaussian. ■

6 Problems

6.1 Previous Core Competency Problems

Problem 1 (2018 Summer Practice, # 11)**.** Suppose that $X_1, \dots, X_n \overset{i.i.d.}{\sim} Ber(\lambda/n)$.

(a) What is the distribution of $\sum_{i=1}^n X_i$.

(b) Compute $\lim_{n\to\infty} \mathbb{P}(\sum_{i=1}^n X_i = k)$, where k is any fixed nonnegative integer, and hence show that $\sum_{i=1}^n X_i$ converges in distribution to a random variable Y.

(c) Compute $\mathbb{E}[Y(Y-1)]$, where Y is as in part (b).

Problem 2 (2018 Summer Practice, # 16)**.** Farmers in the Hudson Valley pack apples into bags of approximately 10 pounds, but due to the variation in apples the actual weight varies. We may model the weight of a bag as uniformly distributed in $[9.5, 10.5]$ and independent of other bags. The farmers load 1200 bags onto a truck with maximal admissible load of 13000 pounds. Find a simple approximation to the probability that the truck is overloaded, expressed in terms of the Normal distribution.

Problem 3 (2018 Summer Practice, # 19)**.** suppose for every $n \geq 1$, A_n is a real symmetric matrix of size $n \times n$, whose eigenvalues $(\lambda_1, \dots, \lambda_n)$ satisfies the following properties:

(i) $\max_{i=1}^n |\lambda_i| \overset{n\to\infty}{\to} 0$.

(ii) $\sum_{i=1}^n \lambda_i^2 = 1$.

Find the asymptotic distribution of $\sum_{i,j=1}^n A_n(i,j) X_i X_j$, where $\{X_i\}_{i\geq 1}$ is a sequence of i.i.d. $N(0,1)$.

Problem 4 (2018 September, # 3)**.** Suppose that, for $n \geq 1$, X_n is a random variable taking values in $\{1/n, 2/n, \dots, n/n\}$ with equal probability $1/n$.

(i) Show that X_n converges in distribution, as $n \to \infty$? What is its weak limit?

(ii) Let $f : [0,1] \to \mathbb{R}$ be defined as $f(x) = x\sin(x)$, for $x \in [0,1]$. Using the above or otherwise, show that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n f\left(\frac{k}{n}\right) = \int_0^1 f(x)\, dx.$$

Problem 5 (2018 September, # 7)**.** Suppose $X_1, \dots, X_n$ are i.i.d. with $\mathbb{P}(X_i = \pm 1) = \frac{1}{2}$. Define

$$Y_i := \prod_{j=1}^i X_j, \qquad \text{for } i = 1, \dots, n.$$

(i) Find the joint distribution of (Y_1, Y_2).

(ii) Derive the limiting distribution of $\frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i$.

Problem 6 (2019, May # 2)**.** Let $Z_1, \dots, Z_n$ be i.i.d. random variables with density f. Suppose that (i) $\mathbb{P}(Z_i > 0) = 1$, and (ii) f is continuous on $[0, \epsilon)$, for some $\epsilon > 0$. Let $\lambda := f(0)$. Let

$$X_n = n \min\{Z_1, \dots, Z_n\}.$$

Show that X_n converges in distribution, and find the limiting distribution.

Problem 7 (2019 May, # 8)**.** Suppose you have a quadratic form $\mathbf{X}_n^T A_n \mathbf{X}_n$, where $\mathbf{X}_n \sim N_n(\mathbf{0}_{n\times 1}, \mathbf{I}_{n\times n})$, and A_n is a symmetric $n \times n$ matrix with 0 on the diagonal. Let $(\lambda_{1,n}, \lambda_{2,n}, \dots, \lambda_{n,n})$ denote the eigenvalues of A_n, and let $\|\lambda\|_{2,n} := \sqrt{\sum_{i=1}^n \lambda_{i,n}^2}$ denote the ℓ_2-norm of the eigenvalues.

(a) If $\lim_{n\to\infty} \frac{\max_{i=1,\dots,n} |\lambda_{i,n}|}{\|\lambda\|_{2,n}} = 0$, show that $T_n := \frac{1}{\|\lambda\|_{2,n}} \mathbf{X}_n^T A_n \mathbf{X}_n \xrightarrow{\text{d}} N(0,1)$ as $n \to \infty$.
[Hint: You may use Lyapunov's[1] CLT. Note that the trace of a square matrix is the sum of its eigenvalues.]

(b) If

$$\lim_{n\to\infty} \frac{\lambda_{1,n}}{\|\lambda\|_{2,n}} = 1,$$

show that $T_n \xrightarrow{\text{d}} \chi_1^2 - 1$.

Problem 8 (2019 May, # 9). Let $Y_n = \prod_{i=1}^n X_i$ where $X_1, \dots, X_n$ are i.i.d. nonnegative non-degenerate random variables with mean $\mathbb{E}(X_i) = 1$. Prove that $Y_n \xrightarrow{\text{P}} 0$ as $n \to \infty$ when: (i) $\mathbb{P}(X_1 = 0) > 0$, and (ii) $\mathbb{P}(X_1 = 0) = 0$.

Problem 9 (2019 May, # 10). Let $f_{X,Y}(x, y)$ be a bivariate density and let $(X_1, Y_1), \dots, (X_N, Y_N)$ be i.i.d. $f_{X,Y}$. Let $w(\cdot)$ be an arbitrary probability density function. Let

$$\hat{f}_X(x) = \frac{1}{N} \sum_{i=1}^N \frac{f_{X,Y}(x, Y_i) w(X_i)}{f_{X,Y}(X_i, Y_i)}.$$

Show that, for any $x \in \mathbb{R}$, $\hat{f}_X(x) \xrightarrow{\text{P}} f_X(x)$, where f_X is the marginal density of X_1.

Problem 10 (2019 September, # 6). Suppose that $X_1, X_2, \dots$ are i.i.d. having an exponential distribution with mean 1. Show that

$$\frac{\max_{1\le k\le n} X_k}{\log n} \xrightarrow{\text{P}} 1 \text{ as } n \to \infty$$

where $\xrightarrow{\text{P}}$ denotes convergence in probability.

Problem 11 (2020 May, # 2). Let $X_1, X_2, \dots, X_n$ denote n independent and identically distributed observations from Uniform$(0, 1)$. We order these observations according to their distance from $x = 0.75$ and call the ordered ones $X^x_{(1)}, X^x_{(2)}, \dots, X^x_{(n)}$. Note that $X^x_{(1)}$ and $X^x_{(n)}$ are, respectively, the closest and farthest observations from $x = 0.75$.

(i) Prove that $X^x_{(1)}$ converges to 0.75 in probability.

(ii) What does $X^x_{(n)}$ converge to in probability? Prove your answer.

Problem 12 (2020 September, # 2). Suppose that $X_1, \dots, X_{2n}$ are i.i.d. $U[0, 1]$. Let $Y_i = X_{2i-1} + X_{2i}$ for $1 \le i \le n$.

(a) Find the limiting distribution of Y_1.

(b) Find the limiting distribution of $\sqrt{n}(2 - Y_{(n)})$ as $n \to \infty$.

Problem 13 (2021 September, # 5). Suppose $\{\xi_i\}_{i\ge0}$ are i.i.d. $\mathcal{N}(0, 1)$ random variables. Find the constant c such that

$$\frac{\max_{1\le i\le n} X_i}{\sqrt{\log(n)}} \xrightarrow{\text{P}} c,$$

for each of the following three cases where $\{X_i\}_{i\ge1}$ is defined.

(i) $X_i = \xi_i$ for $i \ge 1$.

(ii) $X_i = \xi_i + \xi_0$ for $i \ge 1$.

(iii) $X_i = \frac{\xi_i + \xi_{i-1}}{\sqrt{2}}$ for $i \ge 1$.

Problem 14 (2021 September, # 7). Let $X_1, X_2, \dots, X_n \overset{\text{i.i.d.}}{\sim} F$ (F denotes the CDF). Our goal is to estimate $\gamma = F(0) + 2F(1)$. We employ the following estimate

$$\hat{\gamma} = \frac{1}{n} \left(\sum_{i=1}^n \mathbf{1}\{X_i \le 0\} + 2 \sum_{i=1}^n \mathbf{1}\{X_i \le 1\} \right),$$

where $\mathbf{1}\{\cdot\}$ dneotes the indicator function.

[1] **Lyapunov's CLT**: Suppose that $\{Z_1, Z_2, \dots\}$ is a sequence of independent random variables such that Z_i has finite expected value μ_i and variance σ_i^2. Define $s_n^2 := \sum_{i=1}^n \sigma_i^2$. If $\lim_{n\to\infty} \frac{1}{s_n^3} \sum_{i=1}^n \mathbb{E}[|Z_i - \mu_i|^3] = 0$ is satisfied, then $\frac{1}{s_n} \sum_{i=1}^n (Z_i - \mu_i) \xrightarrow{\text{d}} N(0, 1)$.

(i) Calculate $\mathbb{E}[\hat{\gamma}]$.

(ii) What is the limiting distribution of $\sqrt{n}(\hat{\gamma} - \gamma)$? Justify your answer.

Problem 15 (2021 September, # 8). Answer the following questions.

(i) Suppose that $(X_n, Y_n) \xrightarrow{d} \mathcal{N}(0, \Sigma)$ in distribution with $\Sigma = [2, 1; 1, 1]$. What does $(X_n - Y_n)^2$ converge in distribution? Prove your answer.

(ii) Suppose that $(X_n, \sqrt{n}Y_n) \xrightarrow{d} \mathcal{N}(0, \Sigma)$ in distribution with $\Sigma = [2, 1; 1, 1]$. What does $(X_n - Y_n)^2$ converge to in distribution? Prove your answer.

(iii) Let $X_n \xrightarrow{P} 1$. For each X_n, we pick Y_n uniformly at random from the internet $[0, X_n]$. What does Y_n converge to in distribution? Prove your answer.