BUSI 573: Stochastic Models in Operations Management **Spring 2026**

Lecture 1: Limit Theorems

Lecturer: Süleyman Kerimov *Date: January 13, 2026*

Disclaimer: *These notes have adapted ideas from several expositional texts, including work by Sheldon M. Ross and Erol A. Peköz. These notes are not meant to be complete or fully rigorous; some proofs are not given, incomplete, or only outlined, as they are discussed in class.*

1.1 Why do we need measure theory?

It is unfortunate that we are starting this course with an informal example. Consider a circle with a radius of 1 meter. We say that two points (a and b) on the edge of the circle belong to the same family if you can go from a to b, or, b to a, by traveling 1 meter around the edge of the circle. Alternatively, you can consider an equivalence relation on the interval $I = [0, 2\pi)$, where $a, b \in I$ belong to the same equivalence class if the distance from a to b is 1, given that you are allowed to loop the interval.

Now each family will pick one of its members as a representative. What is the probability that a point a selected uniformly at random on the edge of the circle is a representative? At first glance, you may suspect that the answer is probably not 1, maybe it is 0.

Note that each family has infinitely many members: once you start from a point a, you will never visit point a again. This is because the circumference of the circle is 2π, which is irrational. Consider the following events

$$A = \{a \text{ is a representative}\},$$

$$B_i = \{a \text{ is } i \text{ steps clockwise from the representative of its family}\},$$

$$C_i = \{a \text{ is } i \text{ steps counter-clockwise from the representative of its family}\}.$$

Since a is chosen uniformly at random, we must have $\mathbb{P}(A) = \mathbb{P}(B_i) = \mathbb{P}(C_i)$ by symmetry. Moreover, since every family has a representative, we must have

$$\mathbb{P}(A) + \sum_{i=1}^{\infty} (\mathbb{P}(B_i) + \mathbb{P}(C_i)) = 1. \tag{1.1}$$

Let $x = \mathbb{P}(A)$. Then per (1.1), we get $x + \sum_{i=1}^{\infty} 2x = 1$, which has no solution for $x \in [0, 1]$. The event A is an example of a non-measurable event, because we cannot measure its probability. The reason why the example is not completely formal is that choosing exactly one representative from each family requires the axiom of choice, which we will not discuss.

Discussion 1.1. *When $X \sim U[0,1]$, the following looks contradictory:* $1 = \mathbb{P}(0 \leq X \leq 1) = \sum_{x\in[0,1]} \mathbb{P}(X = x) = 0$?

Discussion 1.2. *Argue that the set of rational numbers $\mathbb{Q}$ is countable (Cantor snake), and the set of irrational numbers $\mathbb{R} \setminus \mathbb{Q}$ is uncountable (Cantor's diagonal argument).*

1.2 Probability spaces

Let Ω be an arbitrary set of points ω. For our purposes, Ω consists of all the possible results or outcomes ω of an experiment or observation. Next we define a collection of subsets of Ω, where these subsets can be viewed as events for which we can calculate a probability.

Definition 1.3. *The collection of sets $\mathcal{F}$ is a sigma field (we also say σ-field), if it has the following properties:*

1. $\Omega \in \mathcal{F}$,
2. *If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$,*
3. *If $A_1, A_2, ... \in \mathcal{F}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.*

We note that by DeMorgan's law, which states that $(\cup_{i=1}^{\infty} A_i)^c = \cap_{i=1}^{\infty} A_i^c$, (3) in Definition 1.3 can be replaced with: if $A_1, A_2, ... \in \mathcal{F}$, then $\cap_{i=1}^{\infty} A_i \in \mathcal{F}$. Therefore, σ-algebra is simply a non-empty collection of subsets of Ω, which is closed under countable unions, countable intersections, and complement. $(\Omega, \mathcal{F})$ is also referred as a measurable space.

Definition 1.4. *A probability space is a measure space with total measure one. It is denoted by $(\Omega, \mathcal{F}, \mathbb{P})$, where*

- *Ω is a set (also known as sample space)*
- *$\mathcal{F}$ is a σ-field of subsets of Ω (the sets in $\mathcal{F}$ are also known as events)*
- *$\mathbb{P}$ is a function from $\mathcal{F}$ to $[0,1]$ that satisfies $\mathbb{P}(\Omega) = 1$ and if $A_1, A_2, ... \in \mathcal{F}$ are pairwise disjoint, then*
$$\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

We write $\sigma(\mathcal{A})$ to represent the smallest σ-field that contains the collection of events $\mathcal{A}$. We also say that $\sigma(\mathcal{A})$ is the σ-field generated by $\mathcal{A}$. Let's say we want to calculate probabilities on the sample space $\Omega = [0,1]$ (for example, we want to sample a uniform random number from this interval). One natural candidate for a σ-field $\mathcal{F}$ would be the collection of all possible subsets of Ω. But if you remember our informal example in the introduction, we will not be able to equip this σ-field with a probability measure $\mathbb{P}$, since sets like the set of representatives will belong to $\mathcal{F}$. What is the next natural try? Consider the σ-field generated by the set of all singletons: $\mathcal{F} = \sigma(\{x\}_{x\in[0,1]})$. But now, how can we calculate the probability that if a uniformly sampled random number belongs to the interval $[0, 0.5]$? We cannot represent this interval (which is an uncountable set) with a countable union of singletons. It turns out that the correct σ-field (which is called the Borel σ-field) is the smallest σ-field generated by all intervals of the form $[x, y)$: $\mathcal{B} = \sigma([x, y)_{x<y, x,y\in[0,1]})$. Finally, once you consider the Lebesgue measure, defined by $\mathbb{P}([x, y)) = y - x$ for $0 \le x \le y \le 1$, we are basically good to go.

Discussion 1.5. *Argue that singletons, set of rational and irrational numbers are in the Borel σ-field on* $[0, 1]$.

Next, we discuss the continuity property of the probability function $\mathbb{P}$. Let $(A_n)_{n\ge1}$ be a sequence of events, and let

$$\liminf A_n := \cup_{n=1}^{\infty} \cap_{i=n}^{\infty} A_i,$$
$$\limsup A_n := \cap_{n=1}^{\infty} \cup_{i=n}^{\infty} A_i.$$

Note that by definition, we have $\liminf A_n \subset \limsup A_n$: $\liminf A_n$ consists of all outcomes that are contained in all but a finite number of events $(A_n)_{n\ge1}$, and $\limsup A_n$ consists of all outcomes that are contained in an infinite number of events $(A_n)_{n\ge1}$. We say that $\lim_n A_n$ exists if $\limsup A_n = \liminf A_n$.

We say that $(A_n)_{n\ge1}$ is an increasing sequence of events if $A_n \subset A_{n+1}$ for all $n \ge 1$. Note that $\cap_{i=n}^{\infty} A_i = A_n$ in this case, thus, $\liminf A_n := \cup_{n=1}^{\infty} \cap_{i=n}^{\infty} A_i = \cup_{n=1}^{\infty} A_n$. Also note that $\cup_{i=n}^{\infty} A_i = \cup_{i=1}^{\infty} A_i$. Thus, $\limsup A_n := \cap_{n=1}^{\infty} \cup_{i=n}^{\infty} A_i = \cap_{n=1}^{\infty} \cup_{i=1}^{\infty} A_i = \cup_{n=1}^{\infty} A_n$. Therefore, $\lim_n A_n = \cup_{n=1}^{\infty} A_n$.

We say that $(A_n)_{n\ge1}$ is a decreasing sequence of events if $A_n \supset A_{n+1}$ for all $n \ge 1$. Via similar arguments, it follows that $\lim_n A_n = \cap_{n=1}^{\infty} A_n$ in this case.

Proposition 1.6. *If* $\lim_n A_n = A$, *then* $\lim_n \mathbb{P}(A_n) = \mathbb{P}(A)$.

Proof of Proposition 1.6. First, assume that $(A_n)_{n\ge1}$ is an increasing sequence. Consider the sequence of events

$$B_{n+1} = A_{n+1} \cap A_n^c, \quad \forall n \ge 0,$$

where we define $A_0 = \emptyset$. First, note that B_n's are disjoint and that

$$\bigcup_{i=1}^{n} B_i = A_n \quad \text{and} \quad \bigcup_{i=1}^{\infty} B_i = A.$$

Then we can conclude that

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(B_i) = \lim_n \sum_{i=1}^{n} \mathbb{P}(B_i) = \lim_n \mathbb{P}\left(\bigcup_{i=1}^{n} B_i\right) = \lim_n \mathbb{P}(A_n).$$

The proof when $(A_n)_{n\ge1}$ is a decreasing sequence of events, i.e., $A_n \supset A_{n+1}$ for all $n \ge 1$, is similar (via De Morgan's law). Now we consider the general case. Let

$$C_n = \bigcup_{i=n}^{\infty} A_i.$$

Note that the C_n's are decreasing. Therefore,

$$\lim_n \mathbb{P}(C_n) = \mathbb{P}\Big(\lim_n C_n\Big) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} C_n\right).$$

Now let

$$D_n = \bigcap_{i=n}^{\infty} A_i.$$

Note that the D_n's are increasing. Therefore,

$$\lim_n \mathbb{P}(D_n) = \mathbb{P}\Big(\lim_n D_n\Big) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} D_n\right).$$

Note that

$$D_n = \bigcap_{i=n}^{\infty} A_i \subset A_n \subset \bigcup_{i=n}^{\infty} A_i = C_n,$$

which implies that

$$\mathbb{P}(D_n) \leq \mathbb{P}(A_n) \leq \mathbb{P}(C_n).$$

Since $\lim_n A_n = A$ exists, we have

$$\liminf A_n = \limsup A_n = A,$$

where

$$\lim_n \mathbb{P}(D_n) = \mathbb{P}\Big(\liminf A_n\Big) \quad \text{and} \quad \lim_n \mathbb{P}(C_n) = \mathbb{P}\left(\limsup A_n\right),$$

which concludes the proof. ■

1.3 Intermezzo

To refresh your memory of probability theory, please refer to the notes on Canvas. What follows are some definitions included for completeness of this lecture.

Definition 1.7. *A random variable X is a function that assigns a real number to each outcome in a sample space Ω. Formally, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Then a function $X : \Omega \to \mathbb{R}$ is called a random variable if it satisfies $\{\omega \in \Omega \ : \ X(\omega) \leq x\} \in \mathcal{F}$ for all $x \in \mathbb{R}$. We also say that the random variable X is $\mathcal{F}$-measurable.*

Given a random variable X, we define the σ-algebra generated by X, denoted by $\sigma(X)$, as the smallest σ-algebra with respect to which X is measurable, that is

$$\sigma(X) = \sigma(X^{-1}(B),\, B \in \mathcal{B}(\mathbb{R})) = \{X^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\}.$$

Example 1.8. *Consider an experiment, where we flip two coins, and let X be the number of heads. Note that $X^{-1}(\{0\}) = \{TT\}$, $X^{-1}(\{1\}) = \{HT, TH\}, X^{-1}(\{2\}) = \{HH\}$. The σ-algebra must contain the complements too, so that the σ-algebra generated by X is*

$$\sigma(X) = \{\emptyset, \{TT\}, \{HT, TH\}, \{HH\}, \{HT, TH, HH\}, \{TT, HT, TH\}, \{HH, HT, TH, TT\}, \{HH, TT\}\}.$$

Definition 1.9. *Let X be a random variable and g be any function. If X is discrete, then the expectation of $g(X)$ is defined as*

$$\mathbb{E}[g(X)] = \sum_{x\in\Omega} g(x)f(x),$$

where f is the probability mass function of X. If X is continuous then the expectation of $g(X)$ is defined as

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)\,dx,$$

where f is the probability density function of X.

1.4 Lebesgue's dominated convergence theorem

We are now concerned with fundamental results on interchanging limits and expectations of random variables. We start with some preliminaries.

Definition 1.10. *The sequence of random variables $X_n, n \geq 1$, is said to converge* almost surely *to a random variable X, written $X_n \to_{\text{a.s.}} X$, if*

$$\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1.$$

An equivalent definition is the following. We say that $X_n \to_{\text{a.s.}} X$ if and only if for any $\epsilon > 0$,

$$\lim_{m\to\infty} \mathbb{P}(|X_n - X| < \varepsilon \text{ for all } n \geq m) = 1.$$

Now consider the following example. Let $U \sim U(0,1)$ and $X_n = n\mathbf{1}_{\{n<1/U\}}$. Note that $X_n \to 0$ a.s., and therefore, $\mathbb{E}[\lim_{n\to\infty} X_n] = 0$. On the other hand, $\mathbb{E}[X_n] = n\mathbb{P}(U < 1/n) = 1$ for all n, and therefore, $\lim_{n\to\infty} \mathbb{E}[X_n] = 1$. One should be very careful when interchanging limits and expectations! Lebesgue's Dominated Convergence Theorem is a beautiful theorem that allows us to interchange limits and expectations safely, i.e., it tells us under what condition we can write $X_n \to X$ a.s. $\Rightarrow \mathbb{E}[\lim_{n\to\infty} X_n] = \lim_{n\to\infty} \mathbb{E}[X_n]$.

Theorem 1.11 (Monotone Convergence Theorem)**.** *If a sequence of non-negative random variables increasingly converge to a random variable (written $0 \leq X_n \uparrow X$), then $\mathbb{E}[X_n] \uparrow \mathbb{E}[X]$.*

Theorem 1.11 can be used to prove the following result.

Proposition 1.12 (Fatou's Lemma)**.** *Let Y be a random variable with $\mathbb{E}[|Y|] < \infty$. Then we have*

- *If* $Y \leq X_n$*, then* $\mathbb{E}[\liminf X_n] \leq \liminf \mathbb{E}[X_n]$.
- *If* $Y \geq X_n$*, then* $\mathbb{E}[\limsup X_n] \geq \limsup \mathbb{E}[X_n]$.

And, Proposition 1.12 can be used to prove the following result.

Theorem 1.13 (Lebesgue's Dominated Convergence Theorem)**.** *Assume that* $X_n \to X$ *a.s., and there is a random variable* Y *with* $\mathbb{E}[Y] < \infty$ *such that* $|X_n| < Y$ *for all* n*. Then* $\mathbb{E}[\lim_{n\to\infty} X_n] = \lim_{n\to\infty} \mathbb{E}[X_n]$.

Proof of Theorem 1.13. Note that $|X_n| < Y$ gives $-Y \leq X_n \leq Y$ for all n. Per Proposition 1.12, we have

$$\mathbb{E}[X] = \mathbb{E}[\liminf X_n] \leq \liminf \mathbb{E}[X_n] \leq \limsup \mathbb{E}[X_n] \leq \mathbb{E}[\limsup X_n] = \mathbb{E}[X],$$

Since $\mathbb{E}[X] = \liminf \mathbb{E}[X_n] = \limsup \mathbb{E}[X_n]$, the limit exists and $\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$. ■

Example 1.14. *Using Proposition 1.11, let's prove that if* $X_i \geq 0$ *for all* $i \geq 1$*, then* $\mathbb{E}[\sum_{i=1}^{\infty} X_i] = \sum_{i=1}^{\infty} \mathbb{E}[X_i]$*. We have*

$$\sum_{n=1}^{\infty} \mathbb{E}[X_n] = \lim_{k\to\infty} \sum_{n=1}^{k} \mathbb{E}[X_n] = \lim_{k\to\infty} \mathbb{E}\left[\sum_{n=1}^{k} X_n\right] = \mathbb{E}\left[\sum_{n=1}^{\infty} X_n\right],$$

where the last equality follows from applying the monotone convergence theorem:

$$\sum_{n=1}^{k} X_n \uparrow \sum_{n=1}^{\infty} X_n.$$

The assumption that $X_i \geq 0$ for all $i \geq 1$ is crucial. If we drop this assumption, the statement in Example 1.14 does not hold even if $\sum_{i=1}^{\infty} X_i$ is convergent. Consider $(\alpha_n)_{n=1}^{\infty}$ to be independent and identically distributed (i.i.d.) random variables with $\mathbb{P}(\alpha_1 = \pm 1) = 1/2$, and define a stopping time $\tau = \inf\{n \geq 1 : \sum_{k=1}^{n} \alpha_k = 1\}$. We will cover stopping times later in this lecture, but convince yourself that $\mathbb{P}(\tau < \infty) = 1$. Let $X_n = \alpha_n \mathbf{1}_{\{\tau \geq n\}}$. Then, we have that

$$\sum_{n=1}^{\infty} X_n = \sum_{n=1}^{\infty} \alpha_n \mathbf{1}_{\{\tau \geq n\}} = \alpha_1 + \cdots + \alpha_\tau = 1,$$

so that $\mathbb{E}\left[\sum_{n=1}^{\infty} X_n\right] = 1$.

Since the event $\{\tau \geq n\}$ belongs to $\sigma\{\alpha_1, \ldots, \alpha_{n-1}\}$ (we will discuss more about this later, but this basically means the occurrence of the event $\{\tau \geq n\}$ can be determined on the information available by all realizations $\{\alpha_1, \ldots, \alpha_{n-1}\}$), α_n and $\mathbf{1}_{\{\tau \geq n\}}$ are independent. Thus, we get

$$\mathbb{E}[X_n] = \mathbb{E}[\alpha_n]\,\mathbb{E}[\mathbf{1}_{\{\tau \geq n\}}] = 0, \quad n \geq 1.$$

Thus $\sum_{n=1}^{\infty} \mathbb{E}[X_n] = 0 \neq \mathbb{E}[\sum_{n=1}^{\infty} X_n]$.

1.5 Convergence

Here we discuss two types of convergence: convergence in probability and convergence in distribution. Before that, we present a useful result.

Proposition 1.15 (Borel-Cantelli Lemma). *Let $(A_n)_{n=1}^{\infty}$ be a sequence of events.*

1. *If $\sum_{i=1}^{\infty} \mathbb{P}(A_i) < \infty$, then $\mathbb{P}(\limsup A_n) = 0$.*
2. *If $\sum_{i=1}^{\infty} \mathbb{P}(A_i) = \infty$ and all events are independent, then $\mathbb{P}(\limsup A_n) = 1$.*

Proof of Proposition 1.15. We prove (1) first.

$$\mathbb{P}\left(\limsup A_n\right) = \mathbb{P}\left(\bigcap_{k=1}^{\infty} \bigcup_{j=k}^{\infty} A_j\right) \leq \mathbb{P}\left(\bigcup_{j=k}^{\infty} A_j\right) \quad \text{for any } k \geq 1.$$

Then the result follows since

$$\lim_{k\to\infty} \mathbb{P}\left(\bigcup_{j=k}^{\infty} A_j\right) = 0.$$

Now we prove (2). Let $B = \limsup A_n$. We will show that $\mathbb{P}(B^c) = 0$. Let

$$C_i = \bigcap_{n \geq i} A_n^c.$$

Then we have

$$B^c = \bigcup_{i=1}^{\infty} C_i.$$

Thus, we are done if $\mathbb{P}(C_i) = 0$ for all $i \geq 1$. For each i and $k \geq i$, we have

$$\mathbb{P}(C_i) = \mathbb{P}\left(\bigcap_{n=i}^{\infty} A_n^c\right) \leq \mathbb{P}\left(\bigcap_{n=i}^{k} A_n^c\right) = \prod_{n=i}^{k} (1 - \mathbb{P}(A_n)).$$

Now we utilize the fact that $\log(1-x) \leq -x$ for all $x \in [0,1]$. This implies, for all $k \geq i$,

$$\log(\mathbb{P}(C_i)) \leq \sum_{n=i}^{k} \log(1 - \mathbb{P}(A_n)) \leq -\sum_{n=i}^{k} \mathbb{P}(A_n).$$

If this is true for all $k \geq i$, then

$$\log(\mathbb{P}(C_i)) \leq \lim_{k\to\infty} -\sum_{n=i}^{k} \mathbb{P}(A_n) = -\infty.$$

Hence, $\mathbb{P}(C_i) = 0$ for all $i \geq 1$. Note that (1) implies almost surely, only finitely many A_n's will occur, and (2) implies almost surely, infinitely many A_n's will occur. ■

Discussion 1.16. *Consider the following experiment. We toss a coin every minute. The probability that we get H on minute n is $1/n$. Argue that almost surely, infinitely many heads will occur. If the probability is $1/n^2$, then only finitely many times heads will occur, almost surely.*

Definition 1.17. $(X_n)_{n=1}^{\infty}$ *converges in probability to a random variable X (written $X_n \to_{\mathrm{p}} X$), if for any $\epsilon > 0$, $\mathbb{P}(|X_n - X| > \epsilon) \to 0$ as $n \to \infty$.*

From the statement, it is immediate that almost sure convergence implies convergence in probability. The opposite is not true as the following example shows.

Example 1.18. *Let $(X_n)_{n\geq 1}$ be a sequence of random variables with*

$$\mathbb{P}(X_n = 1) = \frac{1}{n}, \quad \text{and} \quad \mathbb{P}(X_n = 0) = 1 - \frac{1}{n}.$$

Note that for any $\varepsilon > 0$,

$$\mathbb{P}\{|X_n - 0| > \varepsilon\} = \frac{1}{n} \to 0 \quad \text{as } n \to \infty,$$

so that $X_n \to_{\mathrm{p}} 0$. But since $\sum_{n=1}^{\infty} \mathbb{P}(X_n = 1) = \infty$, we have $X_n = 1$ for infinitely many values of n, so we do not have almost sure convergence.

Theorem 1.19. *If $X_n \to_{\mathrm{p}} X$, then there is a subsequence $(X_{n_k})_{k\geq 1}$ which converges to X almost surely.*

Proof of Theorem 1.19. Since for every $\varepsilon > 0$, $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$, we can find an index n_1 such that

$$\mathbb{P}(|X_{n_1} - X| > 1/2) < 1/2.$$

Similarly, we can find an index $n_2 > n_1$ such that

$$\mathbb{P}(|X_{n_2} - X| > 1/4) < 1/4.$$

Repeating the argument above, we get a subsequence $(X_{n_k})_{k\geq 1}$ such that for all $k \geq 1$,

$$\mathbb{P}(|X_{n_k} - X| > 1/2^k) < 1/2^k.$$

Since the series $\sum_{k=1}^{\infty} \mathbb{P}(|X_{n_k} - X| > 2^{-k})$ converges, per Borel-Cantelli Lemma (Proposition 1.15), only finitely many events

$$A_k = \{|X_{n_k} - X| > 2^{-k}\}$$

occur almost surely. Therefore, $X_{n_k} \to X$ almost surely. ∎

Definition 1.20. *Let F_n be the distribution function of X_n, and let F be the distribution function of X. We say that X_n converges in distribution to X if $\lim_{n\to\infty} F_n(x) = F(x)$ for all x at which F is continuous.*

Proposition 1.21. *If $X_n \to_{\mathrm{p}} X$, then $X_n \to_{\mathrm{d}} X$. The converse is not true.*

Example 1.22. *Let $(X_n)_{n\geq 1}$ be a sequence of Bernoulli random variables with $p = 1/2$. Also, let $X \sim \mathrm{Bernoulli}(1/2)$. Then clearly $X_n \to_{\mathrm{d}} X$. But we don't have convergence in probability, since $\mathbb{P}(|X_n - X| \geq \epsilon) = 1/2$ for $\epsilon \in (0, 1)$ and for any $n \geq 1$.*

Example 1.23. *Let $(X_n)_{n\geq 1}$ be a sequence of random variables with*

$$F_n(x) = \begin{cases} 1 - \left(1 - \dfrac{1}{n+1}\right)^{(n+1)x}, & x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Then $X_n \to_{\mathrm{d}} X$, where $X \sim \exp(1)$. Clearly, for $x \leq 0$, $F_n(x) = F_X(x)$. For $x \geq 0$, we also have

$$\lim_{n\to\infty} F_n(x) = 1 - \lim_{n\to\infty} \left(1 - \frac{1}{n+1}\right)^{(n+1)x} = 1 - e^{-x} = F_X(x).$$

1.6 Law of large numbers

We will discuss more about probability inequalities later in the course, but we need some of them now.

Proposition 1.24 (Markov's inequality)**.** *If X is a nonnegative random variable, then for any $a > 0$ we have*

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Proof of Proposition 1.24. Let $\mathbb{1}_{\{X\geq a\}}$ be the indicator function, which is 1 if $X \geq a$, and it is 0 otherwise. Since $X \geq 0$, clearly we have $a\mathbb{1}_{\{X\geq a\}} \leq X$. Taking expectations proves the result. ∎

Discussion 1.25. *Here is a stronger version of Markov's inequality. If X is a nonnegative random variable, then for any $a > 0$ we have*

$$\mathbb{P}(X \geq U \cdot a) \leq \frac{\mathbb{E}[X]}{a},$$

where $U \sim U(0,1)$.

Proposition 1.26 (Chebyshev's inequality)**.** *If X is a random variable with $Var[X] < \infty$, then for any $b > 0$ we have*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq b) \leq \frac{Var(X)}{b^2}.$$

Proof of Proposition 1.26. Since $(X - \mathbb{E}[X])^2$ is a nonnegative random variable, per Markov's inequality with $a = b^2$, we get

$$\mathbb{P}((X - \mathbb{E}[X])^2 \geq b^2) \leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{b^2},$$

$$\Rightarrow \mathbb{P}(|X - \mathbb{E}[X]| \geq b) \leq \frac{Var(X)}{b^2}.$$

∎

Theorem 1.27 (The Weak Law of Large Numbers)**.** *If $(X_i)_{i=1}^{\infty}$ are i.i.d. with $\mu := \mathbb{E}[X_1] < \infty$, then for any $\epsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\Big(\Big|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\Big| > \epsilon\Big) = 0$$

Proof of Theorem 1.27. Note that the expectation of $\frac{1}{n}\sum_{i=1}^{n} X_i$ is μ, and its variance is $\frac{\sigma^2}{n}$. Then per Chebyshev's inequality, we have

$$\lim_{n\to\infty} \mathbb{P}\Big(\Big|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\Big| > \epsilon\Big) \le \lim_{n\to\infty} \frac{\sigma^2}{n\epsilon^2} = 0.$$

■

Theorem 1.28 (The Strong Law of Large Numbers)**.** *If $(X_i)_{i=1}^{\infty}$ are i.i.d. with $\mu = \mathbb{E}[X_1] < \infty$, then*

$$\mathbb{P}\Big(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \mu\Big) = 1.$$

Proof of Theorem 1.28. We will prove a weaker version of the statement, where we assume that $K := \mathbb{E}[X_1^4] < \infty$. Further assume that $\mu = 0$, and we generalize the proof in the end. Let $S_n = \sum_{i=1}^{n} X_i$ and consider

$$\mathbb{E}[S_n^4] = \mathbb{E}\left[(X_1 + \cdots + X_n)(X_1 + \cdots + X_n)(X_1 + \cdots + X_n)(X_1 + \cdots + X_n)\right].$$

Expanding the right-hand side, we get terms of the form

$$X_i^4, \quad X_i^3 X_j, \quad X_i^2 X_j^2, \quad X_i^2 X_j X_k, \quad \text{and} \quad X_i X_j X_k X_l,$$

where $i \neq j \neq k \neq l$. Thanks to our independence assumption, we have

$$\mathbb{E}[X_i^3 X_j] = \mathbb{E}[X_i^3]\mathbb{E}[X_j] = 0,$$

$$\mathbb{E}[X_i^2 X_j X_k] = \mathbb{E}[X_i^2]\mathbb{E}[X_j]\mathbb{E}[X_k] = 0,$$

$$\mathbb{E}[X_i X_j X_k X_l] = 0.$$

For given pair i and j, there will be $\binom{4}{2} = 6$ terms in the expansion in the form of $X_i^2 X_j^2$. Thus, we get

$$\mathbb{E}[S_n^4] = n\mathbb{E}[X_1^4] + 6\binom{n}{2}\mathbb{E}[X_1^2 X_2^2].$$

Using independence again, $\mathbb{E}[S_n^4] = nK + 3n(n-1)\mathbb{E}[X_1^2]^2$. Now, since $0 \le \mathrm{Var}(X_1) = \mathbb{E}[X_1^2] - (\mathbb{E}[X_1])^2$, we have

$$(\mathbb{E}[X_i^2])^2 \le \mathbb{E}[X_i^4] = K.$$

Therefore, we have that

$$\mathbb{E}[S_n^4] \le nK + 3n(n-1)K,$$

which implies that

$$\mathbb{E}\Big[\frac{S_n^4}{n^4}\Big] \leq K/n^3 + 3K/n^2.$$

Therefore, it follows that

$$\mathbb{E}\left[\sum_{n=1}^{\infty}\frac{S_n^4}{n^4}\right] = \sum_{n=1}^{\infty}\mathbb{E}\Big[\frac{S_n^4}{n^4}\Big] < \infty.$$

Now, for any $\epsilon > 0$, it follows from Markov's inequality that

$$\mathbb{P}\Big(\frac{S_n^4}{n^4} > \epsilon\Big) \leq \mathbb{E}\Big[\frac{S_n^4}{n^4}\Big]/\epsilon,$$

and therefore,

$$\sum_{n=1}^{\infty}\mathbb{P}\Big(\frac{S_n^4}{n^4} > \epsilon\Big) < \infty,$$

which implies by the Borel–Cantelli lemma that $S_n^4/n^4 > \epsilon$ for only finitely many n's, almost surely. Since this is true for all $\epsilon > 0$, we can thus conclude that almost surely, we have

$$\lim_{n\to\infty}\frac{S_n^4}{n^4} = 0.$$

If $S_n^4/n^4 \to 0$, then we must also have $S_n/n \to 0$. Hence, we have proven that, almost surely,

$$\frac{S_n}{n} \to 0 \quad \text{as } n \to \infty.$$

For the case when $\mu \neq 0$, we can apply the same arguments to the random variables $X_i - \mu$ to obtain that, almost surely, we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu) = 0.$$

$\blacksquare$