
Probability Comprehensive Exam Notes

Revised: March 18, 2020

Topics Index

1	Sigma algebras and sets	3
1.1	Set theory	3
1.2	Sigma algebras and sigma fields	4
2	Probabilities	7
2.1	Definitions and basic properties	7
2.2	Distribution functions (on \mathbb{R} and \mathbb{R}^n)	8
2.3	Integration review and properties of distribution functions	10
2.4	The product construction	10
3	Independence	12
3.1	Basics and preliminaries	12
3.2	Key examples and results	13
4	Random variables	15
4.1	Definitions and basic properties	15
4.2	Independence	16
4.3	Conditional expectation	16
4.4	Common distributions	17
	More practice problems with other common distributions	21
4.5	Expectation	22
4.6	Variance and covariance	26
4.7	Uniform integrability	26
4.8	Convergence and expectation	27
4.9	Characteristic functions of a random variable	28
	Examples involving characteristic functions	30
4.10	The Levy continuity theorem	31
5	Modes of convergence	32
5.1	Definitions of convergence	32
5.2	Remarks on the equivalence of studying convergence to zero	33
5.3	Flowcharts of implications: convergence hieracrhcly	34
5.4	Convergence theorems (analogos to real analysis)	35
5.5	Properties of the modes of convergence for random variables	36
6	Laws of large numbers	38
6.1	The weak law of large numbers	38
6.2	Large deviations	39
6.3	The strong law of large numbers	40
6.4	Concentration inequalities	42

7	Normal distributions	44
7.1	One-dimensional case	44
7.2	Multidimensional normal distributions	45
8	Central limit theorem	48
8.1	Extra problems with proofs and/or hints	49
9	Useful inequalities and asymptotic estimates	51
9.1	Factorials, binomial coefficients and Stirling's inequality	51
	Inequalities on exponential functions and integrals	51
	Other inequalities	51
10	Misc key facts, results, and other reminders	52

1 Sigma algebras and sets

1.1 Set theory

Definition 1.1. We define the *limit infimum* and *limit supremum* of a sequence of sets by

$$\begin{aligned}\liminf_{n \rightarrow \infty} A_n &= \bigcup_{n \geq 1} \bigcap_{m \geq n} A_m \\ &= \{x : x \text{ is in all but finitely many of the } A_n \text{'s}\} \\ \limsup_{n \rightarrow \infty} A_n &= \bigcap_{n \geq 1} \bigcup_{m \geq n} A_m \\ &= \{x : x \text{ is in infinitely many of the } A_n \text{'s}\} \text{ (}\mathbb{P} \text{ of events } A_n \text{ i.o.)}.\end{aligned}$$

Remark 1.2. Note that

$$\mathbb{P} \left(\limsup_{n \rightarrow \infty} A_n \right) = 0,$$

implies that A_n is true for only finitely many $n \in \mathbb{N}$. In particular, this is useful, for example, if $A_n := \{X \geq n\}$ to show that X cannot be arbitrarily large.

Observe that for any $N \geq 1$:

$$\limsup_{n \rightarrow \infty} A_n \subseteq \bigcup_{i=N}^{\infty} A_i.$$

The following properties are also apparent:

- $\liminf_{n \rightarrow \infty} A_n = (\limsup_{n \rightarrow \infty} A_n^c)^c$;
- $\liminf(A_n) \subseteq \limsup(A_n)$;
- $\left\{ \limsup_{n \rightarrow \infty} \frac{X_m + \dots + X_m}{\sqrt{n}} \right\} \in \sigma(X_m, X_{m+1}, \dots)$;
- $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = \mathbb{P}$ of event i.o..

Theorem 1.3 (Monotonicity of Sequences of Sets and Continuity). *We say that A_n increases to A , written $A_n \nearrow A$, if $A_n \subset A_{n+1}$ for all $n \geq 1$ and $A = \bigcup_{n \geq 1} A_n$. We say that A_n decreases to A , written $A_n \searrow A$, if $A_n \supset A_{n+1}$ for all $n \geq 1$ and $A = \bigcap_{n \geq 1} A_n$. If $A_n \searrow A$, then by continuity we get that*

$$\lim_{n \rightarrow \infty} A_n = \bigcap_{n \geq 1} A_n.$$

For example, if for some sequence of sets $\{E_k\}_{k \geq 1}$, we define $F_n := \bigcup_{m \geq n} E_m$, then we have that $F_n \searrow \bigcap_{n \geq 1} F_n$.

Theorem 1.4 (Continuity of the Lebesgue measure on monotone sets from Real Analysis). *Let $\{E_k\}$ be a sequence of measurable sets in \mathbb{R}^n .*

- (1) *If $E_k \nearrow E$, then $\lim_{k \rightarrow \infty} |E_k| = |E|$;*
- (2) *If $E_k \searrow E$ and $|E_k| < \infty$ for some k , then $\lim_{k \rightarrow \infty} |E_k| = |E|$.*

Proof of 1. If for some $j \geq 1$, $|E_j| = \infty$ then $|E_k| = \infty$ for all $k \geq j$, and hence $|E| = \infty$. Now we can assume that $|E_k| < \infty$ for all $k \geq 1$. We write

$$E = E_1 \cup (E_2 \setminus E_1) \cup \dots = E_1 \cup \bigcup_{k \geq 1} (E_{k+1} \setminus E_k).$$

Then

$$|E| = |E_1| + \sum_{k \geq 1} |E_{k+1} \setminus E_k| = |E_1| + \lim_{n \rightarrow \infty} \sum_{k=1}^n (|E_{k+1}| - |E_k|) = \lim_{n \rightarrow \infty} |E_{n+1}|. \quad \square$$

Proof of 2. We can assume that $|E_1| < \infty$ since otherwise we can truncate the sequence and relabel to obtain the same limit. Moreover, we can write E_1 as a disjoint union of measurable sets:

$$E_1 = E \cup \bigcup_{k \geq 1} (E_k \setminus E_{k+1}).$$

Now we have that

$$\begin{aligned} |E_1| &= |E| + \sum_{k \geq 1} |E_k \setminus E_{k+1}| \\ &= |E| + \lim_{n \rightarrow \infty} \sum_{k=1}^n (|E_k| - |E_{k+1}|) \\ &= |E| + \lim_{n \rightarrow \infty} (|E_1| - |E_{n+1}|), \end{aligned}$$

which implies the result. \square

Proposition 1.5. For X_n, Y_n random variables and any $a \in \mathbb{R}$, $\varepsilon > 0$, we have that

$$\begin{aligned} \limsup\{Y_n > a + \varepsilon\} &\subset \{\limsup Y_n > a\} \subset \{\limsup Y_n > a - \varepsilon\} \\ \limsup\{Y_n < a - \varepsilon\} &\subset \{\limsup Y_n < a\} \subset \{\limsup Y_n < a + \varepsilon\}, \end{aligned}$$

and that

$$\begin{aligned} \limsup_{n \rightarrow \infty} \{X_n > a + \varepsilon\} &\subseteq \left\{ \limsup_{n \rightarrow \infty} X_n > a \right\} \subseteq \limsup_{n \rightarrow \infty} \{X_n > a - \varepsilon\} \\ \limsup_{n \rightarrow \infty} \{X_n < a - \varepsilon\} &\subseteq \left\{ \limsup_{n \rightarrow \infty} X_n < a \right\} \subseteq \limsup_{n \rightarrow \infty} \{X_n < a + \varepsilon\}. \end{aligned}$$

1.2 Sigma algebras and sigma fields

We adopt the conventions that

$$\begin{aligned} \Omega &\equiv \text{sample space} = \text{set of all possible outcomes,} \\ \mathcal{F} &\equiv \text{the collection of all events (subsets of the sample space)} \subseteq \rho(\Omega) \\ T &\equiv \bigcap_{n \geq 1} \sigma(\mathcal{F}_n, \mathcal{F}_{n+1}, \dots) \text{ (the tail field),} \end{aligned}$$

where $\rho(\Omega)$ denotes the *power set of Ω* , or the set of all subsets of Ω .

Definition 1.6 (Algebras and Sigma Fields). The set (of sets) \mathcal{F} is a *algebra* (or *field*) if

- (1) $\Omega \in \mathcal{F}$;

$$(2) A, B \in \mathcal{F} \implies A \cup B \in \mathcal{F};$$

$$(3) A \in \mathcal{F} \implies A^c \in \mathcal{F}.$$

Similarly, the set \mathcal{F} is a σ -algebra (or σ -field) if

$$(1) \Omega \in \mathcal{F};$$

$$(2) A_n \in \mathcal{F} \implies \bigcup_{n \geq 1} A_n \in \mathcal{F};$$

$$(3) A \in \mathcal{F} \implies A^c \in \mathcal{F}.$$

Examples of σ -algebras include the following: $\mathcal{F}_1 = \{\emptyset, \Omega\}$ and $\mathcal{F}_2 = \rho(\Omega)$, the power set of Ω . We have the following consequences of these definitions:

- $A, B \in \mathcal{F} \implies A \cap B \in \mathcal{F}$ since $(A \cap B)^c = A^c \cup B^c$;
- $A, B \in \mathcal{F} \implies A \setminus B \in \mathcal{F}$, since $A \setminus B = A \cap B^c = (A^c \cup B)^c$;
- fields $\subseteq \sigma$ -fields;
- For \mathcal{F} a σ -field, $A_n \in \mathcal{F} \implies \bigcap_{n \geq 1} A_n \in \mathcal{F}$.

Definition 1.7 (λ -Systems and π -Systems). A set $G \subset \rho(\Omega)$ is called a λ -system if

$$(1) \Omega \in G;$$

$$(2) A_n \in G \text{ (disjoint)} \implies \bigcup_{n \geq 1} A_n \in G;$$

$$(3) A, B \in G, A \subset B \implies B \setminus A \in G.$$

If C is a λ -system and in addition, C is closed under intersections (i.e., $A, B \in C \implies A \cap B \in C$), then C is called a π -system.

Definition 1.8 (Notation). Let C be a collection of subsets in Ω . Then $\sigma(C)$ is defined to be the *smallest σ -algebra* containing C :

$$\sigma(C) = \bigcap_{\substack{C \subset \mathcal{F} \\ \mathcal{F} \text{ a } \sigma\text{-algebra}}} \mathcal{F}.$$

Similarly, we define $\lambda(C)$ to be the smallest λ -system containing C .

Theorem 1.9 (Dyukin). *If C is a π -system, then $\lambda(C) = \sigma(C)$.*

Example 1.10 (The Borel Field of \mathbb{R}). Consider the following sets:

$$C_1 = \{(a, b) : a < b\}$$

$$C_2 = \{(a, b] : a < b\}$$

$$C_3 = \{[a, b) : a < b\}$$

$$C_4 = \{(-\infty, a] : a \in \mathbb{R}\}$$

$$C_5 = \{[a, \infty) : a \in \mathbb{R}\}$$

$$C_6 = \{[a, b] : a, b \in \mathbb{Q}\}.$$

Then the *Borel field* $\beta(\mathbb{R}) = \sigma(C_1) = \dots = \sigma(C_6)$. It happens that all of the C_i 's in this example are also examples of π -systems.

Example 1.11 (Spring 2019, #6). Assume $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space such that there exist $X_1, X_2 : \Omega \rightarrow \mathbb{R}$ two independent Bernoulli random variables such that $\mathbb{P}(X_i = 0) = \mathbb{P}(X_i = 1) = 1/2$. Show that Ω must have at least 4 elements. Give an example with Ω having 4 elements together with a sigma algebra such that on it we can define two independent Bernoulli as above. Can you generalize this?

2 Probabilities

2.1 Definitions and basic properties

Throughout we will use the notation that Ω denotes the *sample space* and that \mathcal{F} denotes the σ -field of events. Given a fixed probability \mathbb{P} , we will make reference to the *probability space* $(\omega, \mathcal{F}, \mathbb{P})$.

Definition 2.1. The function $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ is a *probability* if

- $\mathbb{P}(\Omega) = 1$; and
- $(A_n)_{n \geq 1} \subset \mathcal{F}$ (disjoint) $\implies \mathbb{P}[\bigcup_{n \geq 1} A_n] = \sum_{n \geq 1} \mathbb{P}(A_n)$.

The next properties are apparent from the definition of P as a probability function:

- $\mathbb{P}(\emptyset) = 0$;
- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$, $\forall A \in \mathcal{F}$;
- $A \subset B \implies \mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$, since $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$.

Theorem 2.2 (Uniqueness on \mathbb{R}). *If \mathbb{P}, \mathbb{Q} are two probability measures on \mathbb{R} , and if $\mathbb{P}(-\infty, x] = \mathbb{Q}(-\infty, x]$ for all $x \in \mathbb{R}$, then $\mathbb{P}(A) = \mathbb{Q}(A)$ for all $A \in \beta(\mathbb{R})$.*

Definition 2.3 (Continuity and Additivity). A probability \mathbb{P} is *continuous* if $A_n \searrow \emptyset \implies \mathbb{P}(A_n) \rightarrow 0$ as $n \rightarrow \infty$. The function \mathbb{P} possessing *finite additivity* is equivalent to the property that whenever $(A_n)_{n \geq 1}$ are disjoint:

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i).$$

On the other hand, σ -*additivity* is equivalent to

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i),$$

provided that the A_n are disjoint: $A_i \cap A_j = \emptyset$ for all $i \neq j$.

Proposition 2.4. *If \mathbb{P} is finite additive and continuous, then it is σ -additive.*

Example 2.5 (Properties of Probabilities). We have the following additional special properties of probability function \mathbb{P} :

- **Subadditivity:** For any $(A_n) \subseteq \mathcal{F}$, $\mathbb{P}(\bigcup_{n \geq 1} A_n) \leq \sum_{n \geq 1} \mathbb{P}(A_n)$;
- **Fatou:** If $(A_n) \subseteq \mathcal{F}$ is arbitrary, then

$$\mathbb{P}(\liminf A_n) \leq \liminf \mathbb{P}(A_n) \leq \limsup \mathbb{P}(A_n) \leq \mathbb{P}(\limsup A_n).$$

(Compare to Fatou for expectation: If $X_n \geq 0$, then $\mathbb{E}[\liminf X_n] \leq \liminf \mathbb{E}[X_n]$.)

- **Probabilities on countable sets:** If $\Omega = \{\omega_1, \omega_2, \dots\}$ is countable, then for any $A \subseteq \Omega$: $\mathbb{P}(A) = \sum_{\omega_i \in A} \mathbb{P}(\omega_i)$ and $\mathbb{E}[A] = \sum_{\omega_i \in A} \omega_i \cdot \mathbb{P}(\omega_i)$.

2.2 Distribution functions (on \mathbb{R} and \mathbb{R}^n)

Definition 2.6 (Distribution Functions). If μ is a probability on $\beta(\mathbb{R})$, then the *distribution function* of μ is $F_\mu(x) := \mu((-\infty, x])$, or **equivalently**, $F_X(t) = \mathbb{P}[X \leq t]$ for all $t \in \mathbb{R}$. We have the following properties of the distribution function:

- (1) $\lim_{x \rightarrow -\infty} F_\mu(x) = 0$;
- (2) $\lim_{x \rightarrow \infty} F_\mu(x) = 1$;
- (3) F_μ is non-decreasing;
- (4) F_μ is right continuous.

In particular, any function satisfying *all* of these four properties can be taken (shown) to be a distribution function for some random variable.

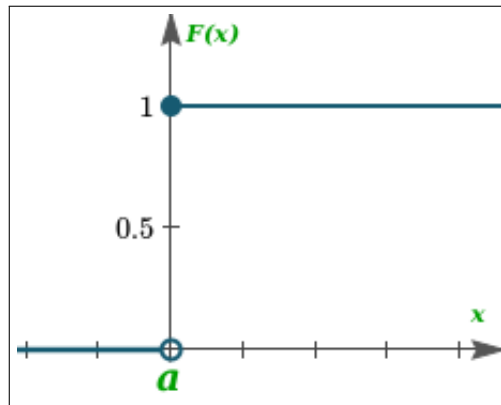
Theorem 2.7. If u, v are probability measures on \mathbb{R} such that $F_u(x) = F_v(x)$ for all $x \in \mathbb{R}$, then $u(A) = v(A)$ for all $A \in \beta(\mathbb{R})$.

Theorem 2.8. Let (Ω, \mathcal{F}) be a probability space with $\mathcal{F} = \sigma(\mathcal{C})$ for \mathcal{C} a π -system. If P, Q are two probabilities on \mathcal{F} such that $P(C) = Q(C)$ for all $C \in \mathcal{C}$, then $P(A) = Q(A)$ for all $A \in \mathcal{F}$.

Example 2.9 (The Dirac Distribution at a Point). For a fixed $a \in \mathbb{R}$, we define

$$\mu(A) := \delta_a(A) = \begin{cases} 1, & a \in A; \\ 0, & a \notin A. \end{cases}$$

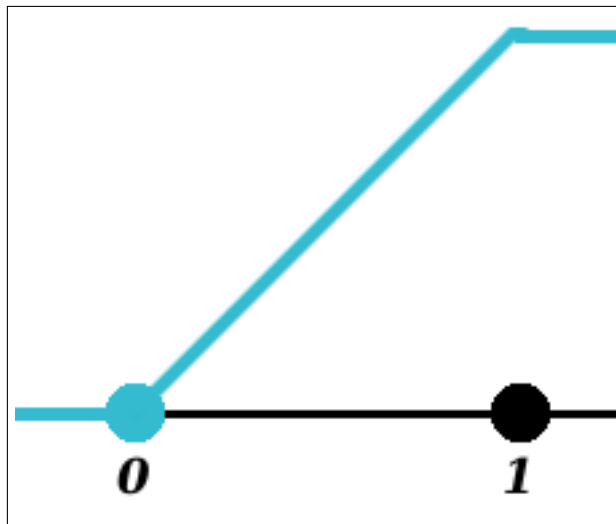
As we can check δ_a is a probability measure on \mathbb{R} since $\delta_a(\mathbb{R}) = 1$ and (A_n) disjoint $\implies \delta_a(\cup_{n \geq 1} A_n) = \sum_{n \geq 1} \delta_a(A_n)$. The distribution function of δ_a , F_{δ_a} , has the following plot shape for arbitrary $a \in \mathbb{R}$:



Example 2.10 (The Lebesgue Measure on the Unit Interval). For an interval $[a, b] \subset [0, 1]$, we define the *Lebesgue measure* by the function $\lambda([a, b]) = b - a$ for $0 \leq a < b \leq 1$ where $\lambda([0, 1]^c) = 0$. Then by integration, we can see that the corresponding distribution function satisfies:

$$F_\lambda(x) = \begin{cases} 0, & x < 0; \\ x, & 0 \leq x < 1; \\ 1, & x \geq 1. \end{cases}$$

This distribution function has the following plot:



Example 2.11 (Construction of μ, F_μ from a Function). Provided that the function F satisfies the four required distribution function properties from the definition above, we can construct a probability measure μ corresponding to this F using the next steps:

- Set $\mu((-\infty, x]) := F(x)$. Then $\mu((a, b]) = \mu((-\infty, b] \setminus (-\infty, a]) = F(b) - F(a)$ for $a < b$.
- To handle the closed intervals, take:

$$\mu([a, b]) = \lim_{n \rightarrow \infty} \mu((a - \frac{1}{n}, b]) = \lim_{n \rightarrow \infty} (F(b) - F(a - 1/n)) = F(b) - F(a^-).$$

- We can apply the *Caratheodory criterion* to see that if μ^* is an outer measure, then

$$\mathcal{A} := \{A \subset \mathbb{R} : \mu^*(B) = \mu^*(B \cap A) + \mu^*(B \setminus A), \forall B \in \beta(\mathbb{R})\},$$

is a σ -field.

- Finally, we can define $\mu := \mu^* \Big|_{\beta(\mathbb{R})}$.

Definition 2.12 (Probabilities on \mathbb{R}^n). Here, for example, we will select $n := 2$ to illustrate the key properties. If μ is a probability on \mathbb{R}^2 , then its distribution function $F_\mu(x, y) := \mu((-\infty, x] \times (-\infty, y])$ satisfies:

- (1) $\lim_{x, y \rightarrow -\infty} F_\mu(x, y) = 0$;
- (2) $\lim_{\substack{x \rightarrow \infty \\ y \rightarrow \infty}} F_\mu(x, y) = 1$;
- (3) F_μ is non-decreasing in each variable;
- (4) F_μ is right continuous in each variable.

Example 2.13 (Spring 2018, #7). Let F_n, F be distribution functions such that $F_n \rightarrow F$ weakly. If F is continuous, show that

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \rightarrow 0.$$

Example 2.14 (Fall 2018, #5). For distribution functions F, G on the real line, define

$$L(F, G) := \inf \{ \varepsilon > 0 : \forall t \in \mathbb{R}, F(t) \leq G(t + \varepsilon) + \varepsilon, G(t) \leq F(t + \varepsilon) + \varepsilon \}.$$

It is known that L is a metric. Prove that $L(F_n, F) \rightarrow 0$ as $n \rightarrow \infty$ if and only if F_n converges weakly to F .

2.3 Integration review and properties of distribution functions

The *mean value theorem* (MVT for derivatives) states that if f is defined and continuous on $[a, b]$ and differentiable on (a, b) , then $\exists c \in (a, b)$ such that $(b - a)f'(c) = f(b) - f(a)$. The *intermediate value theorem* (IVT) states that if f is continuous on $[a, b]$, then for all c in the range between $f(a)$ and $f(b)$, $\exists x \in (a, b)$ such that $f(x) = c$. The *MVT for integrals* states that if f is continuous on $[a, b]$ then $\exists c \in (a, b)$ such that

$$f(c) = \frac{1}{b-a} \int_a^b f(t) dt.$$

Notice that f **monotone increasing** on $[a, b] \implies f \in \text{BV}[a, b]$, i.e., f has *bounded variation* on $[a, b]$. If f has bounded variation on $[a, b]$, then we can express $f = g - h$ where g, h are both bounded and monotone increasing on $[a, b]$. This result is known as the **Jordan decomposition theorem**. Note that these functions can be extended to be monotone increasing on all of \mathbb{R} as well. Also, f monotone increasing implies that f is differentiable a.e., $f' \geq 0$, and $\int_a^b f'(t) dt \leq f(b) - f(a)$ for all $a < b$.

Proposition 2.15 (Monotone increasing function properties). *If f is monotone increasing (especially including distribution functions), then*

- (a) f has at most countably many points of discontinuity.
- (b) f is differentiable a.e.
- (c) f' is a measurable function.
- (d) $f' \in L^1$ and we have the FTC as an upper bound: $0 \leq \int_a^b f' \leq f(b) - f(a)$.

Nested characterizations: We say that f is *Lipschitz on $[a, b]$* if there exists a $K > 0$ (where K is independent of all $x, y \in [a, b]$) such that $|f(x) - f(y)| \leq K \cdot |x - y|$ for all $x, y \in [a, b]$. We have the next important nested inclusion of function types. Here, C^1 denotes the set of functions with one continuous derivative:

$$C^1[a, b] \subsetneq \text{Lip}[a, b] \subsetneq \text{AC}[a, b] \subsetneq \text{BV}[a, b] \subsetneq L^\infty[a, b].$$

Typical counter examples to these types of functions are \sqrt{x} or are of the form $f(x) = x^a \sin(1/x^b)$.

2.4 The product construction

Definition 2.16 (Products). Suppose that $(\Omega_1, \mathcal{F}_1, P_1)$ and $(\Omega_2, \mathcal{F}_2, P_2)$ are two probability spaces. Then the *product construction* between these two spaces corresponds to taking $\Omega := \Omega_1 \times \Omega_2$, $\mathcal{F} := \mathcal{F}_1 \otimes \mathcal{F}_2 := \sigma(\{A \times B : A \in \mathcal{F}_1, B \in \mathcal{F}_2\})$, and the tensor of the probabilities as $\mathbb{P} := P_1 \otimes P_2$, where $\mathbb{P}(A \times B) = P_1(A) \cdot P_2(B)$.

Example 2.17 (Spring 2019, #7). If X, Y are two random variables such that $X \geq Y$ and X, Y have the same distribution, then $X = Y$ almost surely.

Solution due to Michael Lacey. Assume that $X \leq Y$ almost surely. For all y ,

$$\mathbb{P}(Y \leq y) = \mathbb{P}(X \leq Y < y) \leq \mathbb{P}(X \leq y).$$

If $X \neq Y$ a.s., then $\exists y$ so that

$$\mathbb{P}(Y \leq y) < \mathbb{P}(X \leq y).$$

By upper continuity, $\exists y' > y$ so that

$$\mathbb{P}(Y < y') < \mathbb{P}(X \leq y).$$

Thus $\mathbb{P}(X \leq y, Y > y') > 0$, a contradiction. □

3 Independence

3.1 Basics and preliminaries

For our probability space $(\Omega, \mathcal{F}, \mathbb{P})$, in general for any $A, B \in \mathcal{F}$, the conditional probability of A given B has occurred is

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}.$$

As a notion of *independence* of the random variables A and B , we want that observing B does not affect the probability of A , i.e., that $\mathbb{P}[A|B] = \mathbb{P}[A]$ provided that $\mathbb{P}[B] \neq 0$. This is equivalent to the condition that $\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B]$, which holds even when $\mathbb{P}[B] = 0$. Thus we take this property as our definition of independence: $\boxed{\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B]}$.

Example 3.1. First, if $A := \emptyset, \Omega$, then A, B are always independent for any $B \in \mathcal{F}$. We have that if A, B are independent, then A^C, B are independent. We can prove this fact using the following logic:

$$\begin{aligned} \mathbb{P}(A^C \cap B) &= \mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B), \text{ by independence,} \\ &= \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) = (1 - \mathbb{P}(A))\mathbb{P}(B) \\ &= \mathbb{P}(A^C)\mathbb{P}(B). \end{aligned}$$

So to summarize, any of the elements of $\{\emptyset, \Omega, A, A^C\}$ are independent of any in $\{\emptyset, \Omega, B, B^C\}$ whenever A, B are independent.

We generalize to the independence of a set of random variables by defining that A_1, A_2, \dots, A_n are independent if $\forall 1 \leq i_1 < i_2 < \dots < i_k \leq n$ (with $2 \leq k \leq n$) \implies

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \dots \mathbb{P}(A_{i_k}).$$

Note that this definition implies pairwise independence, but is actually a much stronger requirement. For example, A, B, C are independent if each of $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, $\mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C)$, $\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C)$, and $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$. In general, for I some indexing set, we make the next definition: $(\mathcal{F}_i)_{i \in I}$ are independent if $\forall i_1, \dots, i_n \in I$ (distinct) with $A_{i_j} \in \mathcal{F}_{i_j}$ we have

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_n}) = \prod_{j=1}^n \mathbb{P}(A_{i_j}).$$

Theorem 3.2 (Generation Theorem). *For any $i \in I$, assume that $\mathcal{F}_i \in \sigma(C_i)$ for C_i always a π -system. Then $(\mathcal{F}_i)_{i \in I}$ are independent if and only if $\forall i_1, \dots, i_n \in I$ (distinct), and with $A_{i_j} \in C_{i_j}$, then $\mathbb{P}(C_{i_1} \cap \dots \cap C_{i_n}) = \mathbb{P}(C_{i_1}) \dots \mathbb{P}(C_{i_n})$.*

Corollary 3.3 (Grouping Theorem). *If $(\mathcal{F}_i)_{i \in I}$ are independent, and if $\{I_j\}_{j \in J}$ partition I as $I = \dot{\cup}_{j \in J} I_j$, and if $\sigma_j := \sigma(\{\mathcal{F}_i : i \in I_j\})$, then the $(\sigma_j)_{j \in J}$ are independent.*

Example 3.4 (Independent Groupings). Suppose that the sequence of $(A_n)_{n \geq 1}$ are independent. Then for

$$\begin{aligned} B_1 &:= A_1 \cup A_3 \cup A_5 \cup A_7 \cup \dots \\ B_2 &:= (A_2 \cap A_4 \cap A_6) \cup (A_8 \cap A_{10} \cap A_{12}) \cup \dots \\ C_1 &:= A_1 \cup A_4 \cup A_7 \cup A_{10} \cup \dots \\ C_2 &:= (A_2 \cap A_5) \cup (A_8 \setminus A_{11}) \cup \dots \\ C_3 &:= (A_3 \setminus A_6) \cup (A_9 \setminus A_{12}) \cup \dots, \end{aligned}$$

we have that B_1, B_2 and C_1, C_2, C_3 are each independent from the ‘‘grouping’’ corollary above.

Example 3.5 (A Set Independent of Itself). Observe that A is independent of itself (henceforth, a.s. constant per exam 2), then $\mathbb{P}(A) \in \{0, 1\}$. This happens because of our independence requirement that $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2 \implies \mathbb{P}(A) \in \{0, 1\}$.

Example 3.6 (A Three-Variable Proof). If A, B, C are independent, then we can prove also that A^C, B, C are independent:

$$\begin{aligned} \mathbb{P}(A^C \cap B \cap C) &= \mathbb{P}((B \cap C) \setminus A) = \mathbb{P}(B \cap C) - \mathbb{P}(B \cap C \cap A) \\ &= \mathbb{P}(B)\mathbb{P}(C) - \mathbb{P}(B)\mathbb{P}(A)\mathbb{P}(C) = (1 - \mathbb{P}(A))\mathbb{P}(B)\mathbb{P}(C) \\ &= \mathbb{P}(A^C)\mathbb{P}(B)\mathbb{P}(C). \end{aligned}$$

Example 3.7 (Fall 2016, #6). Let $\{A_n\}$ be an infinite collection of independent events. Suppose that $\mathbb{P}(A_n) < 1$ for every $n \geq 1$. Show that $\mathbb{P}(A_n, i.o.) = 1$ if and only if $\mathbb{P}(\cup A_n) = 1$.

3.2 Key examples and results

Theorem 3.8 (Kolmogorov's 0 – 1 Law). Let $(\mathcal{F}_n)_{n \geq 1}$ be an infinite sequence of σ -fields. The corresponding, so-called tail field is

$$T := \bigcap_{n \geq 1} \sigma(\mathcal{F}_n, \mathcal{F}_{n+1}, \dots),$$

where the sets $\sigma(\mathcal{F}_n, \mathcal{F}_{n+1}, \dots)$ are decreasing with n . If the $(\mathcal{F}_n)_{n \geq 1}$ are independent, then $\mathbb{P}(A) \in \{0, 1\}$ for all $A \in T$. As a consequence of this phenomenon, we observe that if $(A_n)_{n \geq 1}$ are **independent**, then

- (1) $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) \in \{0, 1\}$; and
- (2) $\mathbb{P}(\liminf_{n \rightarrow \infty} A_n) \in \{0, 1\}$.

Theorem 3.9 (Borel-Cantelli Lemma (Important, Used Frequently on the Exam)). If $(A_n)_{n \geq 1}$ is such that $\sum_{n \geq 1} \mathbb{P}(A_n) < \infty$, then importantly, $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 0$. This observation is later employed to guarantee almost sure convergence of the A_n to zero provided that the latter sum converges. Secondly, if $(A_n)_{n \geq 1}$ are independent, then

$$\sum_{n \geq 1} \mathbb{P}(A_n) = +\infty \implies \mathbb{P}\left(\limsup_{n \rightarrow \infty} A_n\right) = 1.$$

Proof Note: We can utilize the trick that for any set $A \in \mathbb{F}$, $\mathbb{P}(A^C) = 1 - \mathbb{P}(A) \leq e^{-\mathbb{P}(A)}$, which is essentially the same as saying that $1 - x \leq e^{-x}$ whenever $x \in [0, 1]$. Another useful inequality is that if $|b| \leq 1$, then $|e^b - (1 + b)| \leq |b|^2$.

Example 3.10 (Spring 2019, #2). If $(X_n)_{n \geq 1}$ is a sequence of random variables, then there exists a sequence $(c_n)_{n \geq 1}$ with $c_n \rightarrow \infty$ such that

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \frac{X_n}{c_n} = 0\right) = 1.$$

(**Hint:** Think an easy application of the Borel-Cantelli lemma.)

Example 3.11 (Spring 2018, #8). Let $(X_n)_{n \geq 1}$ be an iid sequence of random variables. Show that $E[X_1^2] < \infty$ if and only if for every $c > 0$, $\mathbb{P}(|X_n| \geq c\sqrt{n} \text{ infinitely often}) = 0$.

The next example gives an interpretation of the condition that $\limsup_{n \rightarrow \infty} A_n = 0$ in terms of probabilities. This technique showed up 2–3 times on the Fall 2019 comprehensive exam.

Example 3.12 (Fall 2019, #2). Assume $(X_n)_{n \geq 1}$ are iid random variables on some space $(\Omega, \mathcal{F}, \mathbb{P})$ with common Gumbel cumulative function given by

$$F(x) = e^{-e^{-x}}, x \in \mathbb{R}.$$

Show that

$$\limsup_{n \rightarrow \infty} (-X_n - \log \log(n)) = 0.$$

Sketch of the general technique. We have to prove two things:

- 1) $\mathbb{P}(\limsup_{n \rightarrow \infty} (-\log \log n - X_n) > \delta) = 0, \forall \delta > 0;$
- 2) $\mathbb{P}(\limsup_{n \rightarrow \infty} (-\log \log n - X_n) > -\delta) = 1, \forall \delta > 0.$

To do this we can utilize the Borel-Cantelli lemma. □

Example 3.13 (Preview of Independent Bernoulli Random Variables). For $n \geq 1$, let $X_n \sim \text{Bernoulli}(p_n)$ so that $\mathbb{P}(X_n = 1) = p_n$ and $\mathbb{P}(X_n = 0) = 1 - p_n$. Then we claim that $X_n \xrightarrow{a.s.} 0 \iff \sum_{n \geq 1} p_n < \infty$. This is true since if the right-hand-side sum converges then

$$\sum_{n \geq 1} \mathbb{P}(X_n = 1) = \sum_{n \geq 1} \mathbb{P}(|X_n - 0| > \varepsilon) < \infty,$$

so that by Borel-Cantelli $\limsup_{n \rightarrow \infty} \mathbb{P}(|X_n| > \varepsilon) = 0$.

Example 3.14. If X_1, \dots, X_n are independent, then for $\phi_1, \dots, \phi_n : \mathbb{R} \rightarrow \mathbb{R}$ measurable, the random variables $Y_i := \phi_i(X_i)$ are still independent. Why is this the case?

$$\{Y_i \in A_i\} = \{\phi_i(X_i) \in A_i\} = \{X_i \in \phi_i^{-1}(A_i)\},$$

where the $\phi_i^{-1}(A_i)$'s are measurable. Hence, the Y_i 's are independent as well.

4 Random variables

4.1 Definitions and basic properties

Definition 4.1. A function $X : \Omega \rightarrow \mathbb{R}$ defined on the sample space is a *random variable* if $X^{-1}(A) \in \mathcal{F}$ for all $A \in \beta(\mathbb{R})$. Here, we define $X^{-1}(A) := \{\omega : X(\omega) \in A\} = \{X \in A\}$ where the second set notation is used more “loosely” as a simplification to use as a nice tool for understanding. We immediately get the following properties which are apparent given our nice new notation for expressing the meaning of a r.v. X :

- (1) $X^{-1}(\bigcup_{i \in I} A_i) = \bigcup_{i \in I} X^{-1}(A_i)$;
- (2) $\{X \in \bigcap_{i \in I} A_i\} = \bigcap_{i \in I} \{X \in A_i\}$; and
- (3) $\{X \in A^C\} = \{X \in A\}^C$, or in other words, $X^{-1}(A^C) = X^{-1}(A)^C$.

For any fixed random variable X , we get a corresponding *probability measure* μ_X defined by

$$\mu_X(B) = \mathbb{P}(X \in B) = \mathbb{P}(X^{-1}(B)), \quad \forall B \in \beta(\mathbb{R}).$$

Generalizations: We can extend this notion more generally by setting $(\Omega, \mathcal{F}) \mapsto (S, \varsigma)$, with \mathcal{F}, ς both σ -fields. In particular, a S -valued random variable is a function $X : \Omega \rightarrow S$ such that $\{X \in A\} \in \mathcal{F}$ for all $A \in \varsigma$. The next couple of results fill in some key details for extended random variables of this type.

Proposition 4.2. *If $X : \Omega \rightarrow S$ is a random variable, and ς a σ -field, then $\mathcal{F}_X = X^{-1}(\varsigma) = \{X^{-1}(A) : A \in \varsigma\}$ is a σ -field. Moreover, if $X : \Omega \rightarrow S$, with $\varsigma = \sigma(\mathcal{C}')$, then $\sigma(X^{-1}(\mathcal{C}')) = X^{-1}(\varsigma) = X^{-1}(\sigma(\mathcal{C}'))$.*

Corollary 4.3 (Characterization for the \mathbb{R} Case). *If $S = \sigma(\mathcal{C}')$, then $X : \Omega \rightarrow S$ is a random variable if and only if $X^{-1}(A) \in \mathcal{F}$ for all $A \in \mathcal{C}'$. In the real-field case, $X : \Omega \rightarrow \mathbb{R}$ is a random variable if and only if $\forall x \in \mathbb{R}$, the sets $\{X \leq x\} \in \mathcal{F}$.*

Proposition 4.4 (Useful General Facts for the Real Case). *We have the next several observations which are collected together here as a matter of reference and completeness:*

- (1) *If X_1, \dots, X_n are random variables on Ω , then the tuple $X \equiv (X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$ is a \mathbb{R}^n -valued random variable;*
- (2) *If $g : \mathbb{R}^n \rightarrow \mathbb{R}$ is a measurable function, i.e., if $g^{-1}(A) \in \beta(\mathbb{R}^n)$ for all $A \in \beta(\mathbb{R})$, then $g(X_1, \dots, X_n)$ is a random variable when the X_i 's are;*
- (3) *If $g : \mathbb{R}^n \rightarrow \mathbb{R}$ is continuous, then g is measurable;*
- (4) *Examples of measurable functions on \mathbb{R}^2 include: $g(x, y) = x + y, x - y, xy, \frac{x}{y}$ ($y \neq 0$), and $g(x, y) = \chi_A(x)\chi_B(y)$ for $A, B \in \beta(\mathbb{R})$ are all measurable, let's say where they are defined.*

Example 4.5 (Limits). If $\{X_n\}_{n \geq 1}$ is any sequence of random variables, then

$$Y := \inf X_n, \sup X_n, \liminf X_n, \limsup X_n,$$

are all random variables.

Example 4.6 (Spring 2017, # 4). Let X, Y be independent, and suppose that each has a uniform distribution on $(0, 1)$. Let $Z := \min(X, Y)$. Find the density $f_Z(z)$ for Z .

4.2 Independence

Definition 4.7 (Independence). If $(X_i)_{i \in I}$ are random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we say that the sequence is independent if $(\mathcal{F}_{x_i})_{i \in I}$ are independent. Here, we write $\mathcal{F}_{x_i} = \{X_i^{-1}(A) : A \in \beta(\mathbb{R})\}$.

The next is a key characterization concerning the independence of random variables: $(X_i)_{i \in I}$ are independent if whenever $i_1, \dots, i_n \in I$ are distinct,

$$\mathbb{P}(X_{i_1} \leq x_{i_1}, \dots, X_{i_n} \leq x_{i_n}) = \mathbb{P}(X_{i_1} \leq x_{i_1}) \cdots \mathbb{P}(X_{i_n} \leq x_{i_n}),$$

for any $x_{i_1}, \dots, x_{i_n} \in \mathbb{R}$.

Proposition 4.8. If X_1, \dots, X_n are independent random variables and g_1, \dots, g_n are Borel measurable functions, then $g_1(X_1), \dots, g_n(X_n)$ are also independent.

Proposition 4.9. If X, Y are independent random variables with both $E[X], E[Y] < \infty$, then $E[|XY|] < \infty$ and $E[XY] = E[X]E[Y]$.

Remark 4.10 (Orthogonal Random Variables). Random variables X, Y such that $E[XY] = E[X]E[Y]$ are called *orthogonal*. It is clear from the above that independent random variables with finite expectation are orthogonal. We also have that if X_1, \dots, X_n are random variables with finite variance, and if X_1, \dots, X_n are pairwise orthogonal, then

$$\text{Var}(X_1 + \cdots + X_n) = \text{Var}(X_1) + \cdots + \text{Var}(X_n).$$

Example 4.11 (Spring 2018, #3). Let X, T be random variables with $E[X], E[Y] < \infty$. Prove that if $E[X|Y] = Y$ and $E[Y|X] = X$, then $X = Y$ a.s.

Example 4.12 (Fall 2018, #2). Suppose X_1, \dots, X_n are iid random variables such that $\mathbb{P}(X_j = +1) = \mathbb{P}(X_j = -1) = 1/2$. Let $S_k := X_1 + \cdots + X_k$ for $1 \leq k \leq n$. Prove that

$$\mathbb{P}\left(\max_{1 \leq k \leq n} S_k \leq \ell\right) = 2\mathbb{P}(S_n > \ell) + \mathbb{P}(S_n = \ell).$$

Example 4.13 (Spring 2019, #8). Assume that X_1, X_2, \dots, X_n are iid with density $f(x) = \frac{2}{x^3}$ for $x \geq 1$ and 0 otherwise. Define

$$M_n := \frac{1}{n} \max\{X_1, \sqrt{2}X_2, \dots, \sqrt{n}X_n\}.$$

Show that M_n converges in distribution and find the limit.

(Hint: Look at the distribution function and take the logarithm of the product where $1 + \frac{1}{2} + \cdots + \frac{1}{n} \sim \log(n) + \gamma$ and $\log(1 - x) \approx -x$ for $0 < x < 1$.)

4.3 Conditional expectation

Most importantly, we must remember the property that

$$\mathbb{E}[\mathbb{E}[X | Y]] = \mathbb{E}[X],$$

and learn where to apply this phenomenon. We also have the following consequences of conditional expectation:

- $\mathbb{E}[Y|Y] = Y$;

- If g, f are Borel functions such that $g(Z), f(X)$ are integrable, then $\mathbb{E}\mathbb{E}[g(Z)f(X)|Z] = \mathbb{E}[g(Z)f(X)]$.

Example 4.14 (Fall 2018, #8). Let X and Y be two independent and positive random variables with respective density F_X, f_Y , and let $g : (0, \infty) \rightarrow (0, \infty)$ be a bounded Borel function. Find

$$E \left[g \left(\frac{X}{Y} \right) \middle| Y \right],$$

the conditional expectation of $g(X/Y)$ given Y , and then infer that $V := X/Y$ has a density that you will identify.

Example 4.15 (Fall 2018, #9). Let X, Y, Z be random variables such that (X, Z) and (Y, Z) are identically distributed. Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a Borel function such that $f(X)$ is integrable.

- Show that $E[f(X)|Z] = E[f(Y)|Z]$, a.s.; and
- Let T_1, \dots, T_n be iid random variables with finite first moment, and let $T = T_1 + \dots + T_n$. Using (i), show that

$$E[T_1|T] = \frac{T}{n}.$$

(Hint: If $(T_n)_{n \geq 1}$ are iid and $\hat{T}_n := T_1 + \dots + T_n$, then each of $(T_1, \hat{T}_n), \dots, (T_n, \hat{T}_n)$ are also iid.)

4.4 Common distributions

Definition 4.16. We introduce models for some of the most common discrete probability distributions below. In each of the following cases, $X : \Omega \rightarrow \mathbb{R}$ (or possibly \mathbb{R}^n) is a random variable, and its probability distribution, or measure function which is what varies from case to case, is denoted by μ_X :

- **Bernoulli:** For a fixed probability $p \in (0, 1)$, we say that $X \sim \text{Bernoulli}(p)$ if $\mathbb{P}(X = 1) = p$ and $\mathbb{P}(X = 0) = 1 - p$. Here, we express the measure function by $\mu_X = p\delta_1 + (1 - p)\delta_0$, where

$$\delta_a(A) = \begin{cases} 1, & a \in A; \\ 0, & a \notin A, \end{cases}$$

is the *Dirac delta function*. Indeed, we can see that this expansion is correct since 1) μ_X is additive (because the δ_i 's are); and 2) it is a probability measure as $\mu_X(\mathbb{R}) = p\delta_1(\mathbb{R}) + (1 - p)\delta_0(\mathbb{R}) = 1$. It is standard information that $E[X] = p$, and $\text{Var}(X) = pq$, where $q = 1 - p$ is the flipped probability for the distribution.

Sums of Bernoulli random variables: Note that a sum of n iid Bernoulli random variables with parameter p have mean $\mu = np$ and variance $np(1 - p)$. In particular, if $(X_n)_{n \geq 1}$ are iid with $X_1 \sim \text{Bernoulli}(p)$, then $S_n := X_1 + \dots + X_n \sim \text{Binomial}(n, p)$.

Flipping the inequality in large deviations: If $X_k \sim \text{Ber}(1/2)$ and $Y_k := 1 - X_k$, then also $Y_k \sim \text{Ber}(1/2)$. Thus

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(S_n \leq an) = -I(1 - a), a < 1/2.$$

- **Uniform:** $X \sim \text{Uniform}$ (on a finite, or countable set) if $S = \{x_1, \dots, x_n\}$ and we have that $\mu_X(\{x_1\}) = \dots = \mu_X(\{x_n\}) = \frac{1}{n}$. Alternately, we can state this condition as $X : \Omega \rightarrow \mathbb{R}$ is uniform on S provided that $\mathbb{P}(X = x_1) = \dots = \mathbb{P}(X = x_n) = \frac{1}{n}$.

- **Geometric:** $X \sim \text{Geom}(p)$, which models

$X := \#$ flips in a sequence of coin tosses until the first H occurs.

We have: $\mathbb{P}(X = k) = p(1 - p)^{k-1}$ for any natural numbers $k \geq 1$.

- **Binomial:** $X \sim \text{Binomial}(n, p)$, where X models $X := \#$ H 's in n coin flips, and is such that $\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$ for $0 \leq k \leq n$. We can show that $E[X] =: \mu = np$, and $\text{Var}(X) = np(1 - p) \equiv npq$. The factorial moments are given by $E[(X)_r] = (n)_r p^r$.
- **Poisson:** $X \sim \text{Poisson}(\lambda)$, for $\lambda > 0$, if for all integers $k \geq 0$: $\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$. The *Poisson distribution* naturally arises as the limit of a binomial distribution when the expected number of successes stays fixed at λ : namely, the limiting distribution of $X \sim \text{Binomial}(n, \lambda/n)$ as $n \rightarrow \infty$. It can be verified that if X is Poisson with parameter $\lambda > 0$, then $E[X] = \lambda$, $E[X(X - 1)] = \lambda^2$, and $\text{Var}(X) = E[X^2] - E[X]^2 = \lambda$. The factorial moments are given by $E[(X)_r] = \lambda^r$.

Sums of Poisson random variables: The sum of Poisson random variables is another Poisson random variable. Namely, if $(X_n)_{n \geq 1}$ are random variables such that $X_i \sim \text{Poisson}(\lambda_i)$, then $S_n := X_1 + \dots + X_n \sim \text{Poisson}(\sum_{i=1}^n \lambda_i)$. A special case is that the sum of n $\text{Poisson}(1)$ random variables has the distribution $\text{Poisson}(n)$, and the corresponding characteristic function $\varphi_{\mathcal{P}(n)}(t) = e^{n(e^t - 1)}$ for $t \in \mathbb{R}$.

Example 4.17 (Preview of Continuous Random Variable Distributions). X is a *continuous random variable* if its probability measure, μ_X , has a consistent density function. Namely, f satisfying $f \geq 0$, $\int_{\mathbb{R}} f = 1$, and where $\mu_X(A) = \int_A f(x) dx$. The following listing provides a brief overview of the properties of several of the most common cases we will see:

- **Exponential:** With the construction of this model distribution, we have that the density function is non-zero, and equal to $f(x) = \lambda e^{-\lambda x}$, iff and only if $x > 0$. It is defined to be zero otherwise. Here, to ensure that we get a resulting probability measure, the requirement is that $\lambda > 0$, for example as in other instances in the next sections, with $\lambda := 1$. Note that since we have that $\mathbb{P}(X \geq t) = e^{-\lambda t}$, $\forall t \geq 0$, we can compute that

$$\begin{aligned} E[X] &= \int_0^{\infty} \lambda x e^{-\lambda x} dx = \frac{1}{\lambda} \\ E[X^2] &= \frac{2}{\lambda^2} \\ \text{Var}(X) &= E[X^2] - E[X]^2 = \frac{1}{\lambda^2} \\ E[X^n] &= \frac{n!}{\lambda^n}. \end{aligned}$$

The exponential distribution has the so-called *loss-of-memory property*, which is to say that for all $s, t \geq 0$, $\mathbb{P}(X \geq s + t | X \geq s) = \mathbb{P}(X \geq t)$.

Sums of exponential random variables: Suppose that $(Y_k)_{k \geq 1}$ are iid with $Y_1 \sim \text{Exp}(\lambda)$. Then the CDF of the sum $S_n := Y_1 + \dots + Y_n$ is given by

$$\mathbb{P}(Y_1 + \dots + Y_n \leq t) = 1 - e^{-\lambda t} \sum_{k=1}^{n-1} \frac{(\lambda t)^k}{k!}, t \in \mathbb{R}_{\geq 0}.$$

Suppose that $X_i \sim \text{Exp}(\lambda_i)$ for $\lambda_i > 0$ and $i = 1, 2$. Then if $Z := X_1 + X_2$, the PDF of Z is given by (for $x \geq 0$)

$$f_Z(x) = \begin{cases} \frac{\lambda_1 \lambda_2}{\lambda_2 - \lambda_1} (e^{-\lambda_2 x} - e^{-\lambda_1 x}), & \lambda_1 \neq \lambda_2; \\ \lambda^2 x e^{-\lambda x}, & \lambda_1 = \lambda_2 = \lambda. \end{cases}$$

- **Uniform:** A.k.a., the non-discretized version of a uniformly distributed random variable on some $[a, b] \subset \mathbb{R}$. Then the density and cumulative density functions respectively have the formulas:

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b]; \\ 0, & \text{otherwise,} \end{cases}$$

$$F(x) = \begin{cases} 0, & x < a; \\ \frac{x-a}{b-a}, & x \in [a, b]; \\ 1, & x \geq b. \end{cases}$$

Here, we have that $\mu = \frac{1}{2}(a + b)$ and $\sigma^2 = \frac{(b-a)^2}{12}$, where the density function can be rewritten as

$$f(x) = \begin{cases} \frac{1}{2\sigma\sqrt{3}}, & -\sigma\sqrt{3} \leq x - \mu \leq \sigma\sqrt{3} \\ 0, & \text{otherwise.} \end{cases}$$

We also have that the n^{th} moments are given by

$$E[X^n] = \frac{1}{n+1} \sum_{k=0}^n a^k b^{n-k}.$$

- **Normal and Standard Normal:** $f(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}$. See Section 7 for a substantially complete treatment. Note that $X_i \sim N(0, 1) \implies Y := \sum_{i=1}^n \alpha_i X_i \sim N(0, \alpha_1^2 + \dots + \alpha_n^2)$. If $X_i \sim N(\mu_j, \sigma_j^2)$, then $Y := \sum_{i=1}^n X_i \sim N(\mu_1 + \dots + \mu_n, \sigma_1^2 + \dots + \sigma_n^2)$. If $X_i \sim N(0, 1)$, then $Z_n = X_1 + \dots + X_n$ has the same distribution as $\sqrt{n}Z$. If $Y \sim N(\mu, \sigma^2)$ and $Z \sim N(0, 1)$, then $Y \sim \sigma Z + \mu$ and $Z \sim \frac{Y-\mu}{\sigma}$.

- **Gamma:** For parameters $\lambda, \alpha > 0$, we say that $X \sim \text{Gamma}(\lambda, \alpha)$ if $f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$ for $x > 0$. Here, the *gamma function* is defined by the usual integral $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$. We have the following properties of composite (sum) distributions which result in gamma-distributed random variables:

(G.0) If $\alpha = 1$ then $\text{Gamma}(\lambda, 1) \equiv \text{Exp}(\lambda)$ is the exponential distribution;

(G.1) If X_1, \dots, X_n are iid with $X_1 \sim \text{Exp}(\lambda)$, then $X_1 + \dots + X_n \sim \text{Gamma}(\lambda, n)$;

(G.2) More generally, if X_1, \dots, X_n are iid random variables with $X_i \sim \text{Gamma}(\lambda, \alpha_i)$, then $X_1 + \dots + X_n \sim \text{Gamma}(\lambda, \alpha_1 + \dots + \alpha_n)$;

(G.3) If $Z \sim N(0, 1)$ is standard normal, then $Z^2 \sim \text{Gamma}(\frac{1}{2}, \frac{1}{2})$;

(G.4) If $n \geq 1$ is an integer then the distribution $\text{Gamma}(\frac{1}{2}, \frac{n}{2})$ is called the χ^2 -distribution with n degrees of freedom, which also corresponds to the distribution of $Z_1^2 + \dots + Z_n^2$ where Z_1, \dots, Z_n are iid standard normal random variables.

- **Product distribution:** For two random variables X, Y let $Z := XY$. Then the respective PDF and CDF of Z are given by

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y\left(\frac{z}{x}\right) \frac{dx}{|x|}$$

$$F_Z(z) = \int_0^\infty f_X(x) \int_{-\infty}^{z/x} f_Y(y) dy dx + \int_{-\infty}^0 f_X(x) \int_{z/x}^\infty f_Y(y) dy dx.$$

If X, Y are both square integrable, then

$$\text{Var}(XY) = (\sigma_X^2 + \mu_X^2)(\sigma_Y^2 + \mu_Y^2) - \mu_X^2 \mu_Y^2.$$

The characteristic function of Z is given by

$$\phi_Z(t) = \mathbb{E}_Y[\phi_X(tY)] = \mathbb{E}_X[\phi_Y(tX)].$$

Examples and exercises on exponential random variables

Exercise 4.18. Show that if $X \sim \text{Exp}(\lambda = 1)$ and $U := e^{-X}$ that $U \sim \text{Uniform}(0, 1)$, i.e., prove that the two random variables U, e^{-X} have the same distribution.

Solution to Exercise (I) – Compute matching moments of each rv. We know by a fact that the n^{th} moment of the uniform distribution on $[a, b]$ is given by the following equation for $Y \sim \text{Uniform}(a, b)$:

$$E[Y^n] = \frac{1}{n+1} \sum_{k=0}^n a^k b^{n-k}.$$

This implies that the n^{th} moment for $n \geq 1$ of $\text{Uniform}(0, 1)$ is given by $\frac{1}{n+1}$. Thus it suffices to show that the random variable U defined above has n^{th} moments precisely equal to $\frac{1}{n+1}$ for all $n \geq 1$. To find the expectations $E[U^n] = E[e^{-nX}]$, we can use the composition formula which states that $E[\phi(X)] = \int \phi(x)\mu_X(dx)$. In particular, we have that

$$\begin{aligned} E[U^n] &= \int_0^\infty \lambda e^{-\lambda x} e^{-nx} dx = \int_0^\infty \lambda e^{-(\lambda+n)x} dx \\ &= -\frac{\lambda}{n+\lambda} e^{-(n+\lambda)x} \Big|_0^\infty = \frac{\lambda}{n+\lambda} \mapsto \frac{1}{n+1}, \end{aligned}$$

when $\lambda := 1$. □

Remark 4.19. Note that if the integer-valued moments $\mathbb{E}[X^m]$ for $m \geq 0$ of two random variables each match, then so too must the characteristic functions of each rv for all $t \in \mathbb{R}$, and also the moment generating functions (MGFs), $M(t)$, must match for t where it converges. This follows from the fact that the characteristic function of a rv X , $\varphi_X(t)$, has a Taylor series expansion in it about 0 with

$$\mathbb{E}[X^j] = \frac{\varphi_X^{(j)}(0)}{i^j}, \forall j \geq 0.$$

Solution to Exercise (II) – Compare characteristic functions of each rv. We can compute the characteristic functions of each of $X \sim \text{Exp}(1)$ and e^{-X} and show that they are the same pointwise for all $t \in \mathbb{R}$. Thus by Lévy's continuity theorem that they have the same distributions. Let's do the computations, including integrating by parts to compute $\varphi_U(t)$:

$$\begin{aligned} \varphi_X(t) &= \mathbb{E}e^{itX} = \int_0^\infty e^{itx} e^{-x} dx \\ &= \frac{e^{it} - 1}{it} \\ \varphi_U(t) &= \mathbb{E}e^{ite^{-X}} = \int_0^\infty e^{ite^{-x}} e^{-x} dx \\ &= \int_0^1 e^{-itv} dv \quad (v \mapsto 1 - e^{-x}) \\ &= \frac{e^{it} - 1}{it}. \quad \square \end{aligned}$$

Solution to Exercise (III) – Compute the respective distribution functions of each rv. First, we have that for $u \in [0, 1]$, the distribution function of e^{-U} is given by

$$F_{e^{-U}}(u) = \mathbb{P}(e^{-U} \leq u) = F_{-U}(\log u) = 1 - \log u.$$

Now since $-u \in [-1, 0]$, under the change of variables $x := \log u \mapsto -u = e^{-x}$, we have that $x \in [0, \infty]$. Thus we obtain that

$$F_{e^{-U}}(u) = \begin{cases} 1 - e^{-u}, & u \geq 0, \\ 0, & u < 0. \end{cases}$$

In other words, the distribution functions match for all $u \in \mathbb{R}$, and so the two random variables have the same distribution. \square

Exercise 4.20. For $\mu, \lambda > 0$, suppose that S, T are independent rvs with $S \sim \text{Exp}(\lambda)$ and $T \sim \text{Exp}(\mu)$. Set $V := \min(S, T)$. Prove that **(a)** $V \sim \text{Exp}(\lambda + \mu)$; and **(b)** $\mathbb{P}(V = S) = \frac{\lambda}{\lambda + \mu}$.

Exercise 4.21. For fixed $n \geq 1$ and $1 \leq i \leq n$, suppose that $T_i \sim \lambda_i$ for $\lambda_i > 0$. Set $V_n := \min(T_1, T_2, \dots, T_n)$, and let I denote the (random) index of the T_i 's which is smallest. Prove that **(a)** $\mathbb{P}(V > t) = \exp(-(\lambda_1 + \dots + \lambda_n)t)$; **(b)** $\mathbb{P}(I = i) = \frac{\lambda_i}{\lambda_1 + \dots + \lambda_n}$; and **(c)** that I and V are independent.

Example 4.22 (Spring 2019, #5). If X_1, X_2, \dots, X_n are iid exponential random variables with parameter 1, compute the almost sure limit of

$$\frac{1}{n} \sum_{k=1}^n e^{-X_k - 2X_{k+1} - 3X_{k+2}},$$

as n tends to infinity.

More practice problems with other common distributions

Example 4.23 (Spring 2017, #8). Assume X_1, X_2, \dots are iid standard normal random variables. Show that for any $\lambda > 1/2$,

$$\frac{1}{n^\lambda} (X_1 + \dots + X_n) \xrightarrow{a.s.} 0.$$

Example 4.24 (Fall 2016, #2). Let X be a random variable with continuous density function f and $f(0) > 0$. Let Y be a random variable with

$$Y = \begin{cases} \frac{1}{X}, & \text{if } X > 0; \\ 0, & \text{otherwise,} \end{cases}$$

and let Y_1, Y_2, \dots be iid with distribution equal to that of Y . What is the value of the almost sure limit:

$$\lim_{n \rightarrow \infty} \frac{Y_1 + \dots + Y_n}{n}?$$

(See the second part of this argument.)

Example 4.25 (Spring 2016, #3). Let X_1, X_2, \dots be iid uniform random variables on $(0, 1)$. Show that

$$(X_1 \cdots X_n)^{1/n},$$

converges almost surely as $n \rightarrow \infty$, and compute this limit.

4.5 Expectation

Definition 4.26. If $X = \chi_A$, then $E[X] = \mathbb{P}(A)$. We also use the notation $E[X, A] := E[X\chi_A]$ to denote the expectation of X restricted to A . Any random variable X is called *simple* if $X = \sum_{i=1}^N a_i \chi_{A_i}$ where the A_i 's are disjoint. This implies that $E[X] = \sum_{i=1}^N a_i \mathbb{P}(A_i)$. If $X \geq 0$, then $E[X] = \int X d\mathbb{P}$. If X is an arbitrary random variable, then $X = X_+ - X_-$ where $X_+ = \max(X, 0)$ and $X_- = \max(-X, 0)$. Here we have that $|X| = X_+ + X_-$ (cf. total variation of a function). We say that X is *integrable* if $E[X_+] < \infty$ and $E[X_-] < \infty$.

Proposition 4.27 (Tail-Sum Formula for Expectation). *If $Y \geq 0$ is a continuous random variable, then*

$$\mathbb{E}[Y] = \int_0^\infty \mathbb{P}[Y \geq t] dt.$$

Lemma 4.28 (Necessary Condition for Finiteness of Expectation). *Suppose that X is a random variable. Then*

$$\mathbb{E}[|X|] < \infty \iff \sum_{n \geq 1} \mathbb{P}[X \geq n] < \infty.$$

We make the following observations:

- Note that for X simple, $E[X]$ does not depend on the representation of X . That is, if $X = \sum_i a_i \chi_{A_i} = \sum_j b_j \chi_{B_j}$, then $\sum_i a_i \mathbb{P}(A_i) = \sum_j b_j \mathbb{P}(B_j)$.
- If $X \geq 0$, \exists a sequence $(X_n)_{n \geq 1}$ of simple random variables such that $X_n \nearrow X$ and

$$\lim_{n \rightarrow \infty} E[X_n] = E[X],$$

where this limit is possibly infinite (cf. the monotone convergence theorem).

- To define a general random variable $X \geq 0$ from simple functions, use a limiting procedure: $X_n := \sum_{k=1}^{n^2} \frac{(k-1)}{n} \chi_{\frac{k-1}{n} \leq X \leq \frac{k}{n}}$, so that $X_n \nearrow X$ and $\lim_{n \rightarrow \infty} X_n = X$.

Expectation of a Discrete Random Variable: Suppose that $X : \Omega \rightarrow \{a_1, a_2, a_3, \dots\}$ is a *discrete random variable*. Then

$$E[X] = \sum_{j \geq 1} a_j \cdot \mathbb{P}(X = a_j).$$

Proposition 4.29. *The random variable is integrable $\iff E[|X|] < \infty$, i.e., X is integrable if and only if $|X|$ is integrable.*

Proposition 4.30 (Properties of Expectation). *We have the following key properties:*

- (1) $E[X] \geq 0$ if $X \geq 0$.
- (2.1) If $X \geq 0$, then $E[X] = 0 \iff X = 0$ almost surely.
- (2.2) If X is any random variable taking values on $(-\infty, \infty)$ such that X is square integrable. Then $X = 0$ a.s. $\iff \mathbb{E}[X] = \mathbb{E}[X^2] = 0$.
- (3) $E[\alpha X + \beta Y] = \alpha E[X] + \beta E[Y]$.
- (4) (Monotone convergence theorem) If $X_n \geq 0$ and $X_n \nearrow X$, then $E[X_n] \nearrow E[X]$

(5) $|E[X]| \leq E[|X|]$.

(6) If X, Y are independent and integrable, then XY is integrable, and $E[XY] = E[X]E[Y]$.

(7.1) If X, Y are such that for all bounded, smooth functions $\phi, \psi : \mathbb{R} \rightarrow \mathbb{R}$, $E[\phi(X)\psi(Y)] = E[\phi(X)]E[\psi(Y)]$, then X, Y are independent.

(7.2) $E[XY]^2 \leq E[X^2]E[Y^2]$.

(8) (Fatou's Lemma) If $X_n \geq 0$, then $E[\liminf_{n \rightarrow \infty} X_n] \leq \liminf_{n \rightarrow \infty} E[X_n]$.

(9) (Dominated Convergence Theorem) If $|X_n| \leq Y$ for all $n \geq 1$ with Y integrable, and if $X_n \xrightarrow{a.s.} X$, then $E[X_n] \rightarrow E[X]$.

(10) The r^{th} moments are given by a sum over the factorial moments and the Stirling numbers of the second kind as

$$E[X^r] = \sum_{j=0}^r \left\{ \begin{matrix} r \\ j \end{matrix} \right\} E[(X)_r].$$

Theorem 4.31. If X has distribution μ_X on \mathbb{R} , then

$$E[X] = \int_{\mathbb{R}} x \mu_X(dx).$$

Thus to compute $E[X]$ we only need to know the distribution of X . Also, $E[\phi(X)] = \int \phi(x) \mu_X(dx)$ for any Borel measurable function ϕ . That is, if $f_X(x)$ is the distribution function of X , then $E[\phi(X)] = \int_{-\infty}^{\infty} f_X(x) \phi(x) dx$.

Proposition 4.32 (Key Lemmas and Inequalities). We have the following key named lemmas and inequalities:

(A) (Fubini's Theorem) Let $(\Omega, \mathcal{F}, \mathbb{P}) := (\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, P_1 \otimes P_2)$. Suppose that $F : \Omega \rightarrow \mathbb{R}$ is integrable with $F_{\omega_1}(\omega_2) := F(\omega_1, \omega_2)$. Then

$$\begin{aligned} \int \int F(w_1, w_2) P_1(dw_1) P_2(dw_2) &= \int \left(\int F_{w_1}(w_2) P_2(dw_2) \right) P_1(dw_1) \\ &= \int \left(\int F_{w_2}(w_1) P_1(dw_1) \right) P_2(dw_2). \end{aligned}$$

(B) (Markov's Inequality) If $X \geq 0$, then $\mathbb{P}(X \geq \lambda) \leq \frac{1}{\lambda} E[X]$. Indeed, we see that

$$\begin{aligned} \mathbb{P}(X \geq \lambda) &= \mathbb{P}(\{X \geq \lambda\}) = E[\mathbf{X}_{X \geq \lambda}] \\ &= E\left[\frac{x}{\lambda} \mathbf{X}_{X \geq \lambda}\right] \leq E\left[\frac{x}{\lambda} \mathbf{X}_{X \geq \lambda}\right] \\ &= \frac{1}{\lambda} E[x \cdot \mathbf{X}_{X \geq \lambda}] \leq \frac{1}{\lambda} E[X]. \end{aligned}$$

That is, if $X \geq 0$ is integrable, then the tail probability $\mathbb{P}(X \geq \lambda)$ decays at least as $1/\lambda$ for λ large.

(C) (Chebyshev) If X^2 is integrable and $\lambda > 0$, $\mathbb{P}(|X - E[X]| \geq \lambda) \leq \frac{\text{Var}(X)}{\lambda^2}$. The moral here is that if X^2 is integrable, then the tail probability $\mathbb{P}(X \geq \lambda)$ decays like $1/\lambda^2$ for large $\lambda > 0$.

(D) If $X \in L^p$, then $\mathbb{P}(|X| \geq \lambda) \leq \frac{C}{\lambda^p}$. Indeed, since $\forall \lambda > 0$: $\mathbb{P}(|X| \geq \lambda) = \mathbb{P}(|X|^p \geq \lambda^p) \leq \lambda^{-p} \cdot E[|X|^p]$.

(E) If $E[e^{\alpha|X|}] < \infty$ for some $\alpha > 0$, then $\mathbb{P}(|X| \geq \lambda) \leq Ce^{-\alpha\lambda}$ for all $\lambda > 0$.

(F) (Jensen's Inequality) Let $X = (X_1, \dots, X_n) : \Omega \rightarrow D$ with $D \subset \mathbb{R}^n$ a convex set. If $\phi : D \rightarrow \mathbb{R}$ is convex such that $\phi(X)$ is integrable, and if X is integrable, then $E[\phi(X)] \geq \phi(E[X])$.

Special cases: The function $\phi(x) = |x|^p$ is convex for $p \geq 1$. Then by Jensen's inequality: $E[|X|^p] \geq |E[X]|^p$. Other convex functions include: $\phi(x) := |x|, e^x, -\log x, \frac{1}{x}$.

(G) (Portmanteau's theorem) Suppose that $|X_n - X| \rightarrow 0$ **in probability** (and hence in distribution). For any bounded continuous $f : \mathbb{R} \rightarrow \mathbb{R}$, one has that $E[f(|X_n - X|)] \rightarrow f(0)$.

Proposition 4.33 (Generalized Chebyshev Inequality). Let $f : [0, \infty) \rightarrow [0, \infty)$ be a non-decreasing Borel function, and let $X \geq 0$ be a non-negative random variable. Then for all $a > 0$,

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

provided that $f(X)$ is integrable.

Example 4.34 (Other forms of the Chebyshev inequality). Provided that $\text{Var}(X) = \sigma^2 < \infty$, with $\mathbb{E}X = \mu$ and any $k\varepsilon > 0$, we have that

$$\mathbb{P}(|x - \mu| \geq k\sigma) \leq \frac{1}{k^2} \tag{4.1a}$$

$$\mathbb{P}(|x - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2} \tag{4.1b}$$

$$\mathbb{P}(|x - \mu| < \varepsilon) > 1 - \frac{\sigma^2}{\varepsilon^2}. \tag{4.1c}$$

Proposition 4.35 (Second moment method). If X has finite variance and $X \geq 0$, then

$$\mathbb{P}(X > 0) \geq \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}.$$

If there are positive constants $c_1, c_2 > 0$ such that $E[X^2] \leq c_1 E[X]^2$ and $E[X] \geq c_2$, then $\forall \vartheta \in (0, 1)$:

$$\mathbb{P}(X \geq c_2\vartheta) \geq \frac{(1 - \vartheta)^2}{c_1}.$$

Example 4.36 (Spring 2019, #1). Let X be a non-negative random variable such that $0 < E[X] < +\infty$, and let $0 < x < 1$. Show that

$$\mathbb{P}(X \geq xE[X]) \geq (1 - x)^2 \frac{E[X]^2}{E[X^2]}.$$

Example 4.37 (Spring 2015, #1). Assume X is a symmetric random variable such that $E[X^2] = 1$ and $E[X^4] = 2$. Show that

$$\mathbb{P}(X \geq 1) \leq \frac{14}{27}.$$

Example 4.38 (Spring 2015, #5). for a sequence, X_1, X_2, \dots , of random variables, suppose that we know that

$$\sum_{n \geq 1} nE[|X_n|] < \infty.$$

Show that the sequence $Y_n = X_n + X_{n+1} + \dots + X_{10n}$ converges almost surely and in L^1 to 0. (See for uses of Borel-Cantelli to guarantee a.s. convergence.)

Example 4.39 (Fall 2019, #10). Let X, Y, Z be independent and uniformly distributed on $[0, 1]$. Compute $\mathbb{E}[\min(X, Y, Z)]$.

Theorem 4.40. If $X \geq 0$, then

$$E[X] = \int_0^\infty \mathbb{P}(X \geq \lambda) d\lambda.$$

Corollary 4.41. If $\mathbb{P}(X > \lambda) \leq \frac{C}{\lambda^{1+\varepsilon}}$ for $\varepsilon > 0$, then X is integrable.

Proof. We have that

$$\begin{aligned} E[X] &= \int_0^\infty \mathbb{P}(X > \lambda) d\lambda \\ &\leq \int_1^\infty \frac{C}{\lambda^{1+\varepsilon}} d\lambda + \int_0^1 \mathbb{P}(X \geq \lambda) d\lambda \\ &\leq \frac{C}{\varepsilon} + 1 < \infty. \end{aligned}$$

□

Example 4.42. Let $X \sim \text{Cauchy}$ so that $f_X(x) = \frac{1}{\pi(1+x^2)}$ for $x \in \mathbb{R}$. Then

$$\begin{aligned} \mathbb{P}(|X| \geq \lambda) &= \frac{1}{\pi} \int_{-\infty}^{-\lambda} \frac{dx}{1+x^2} + \frac{1}{\pi} \int_{\lambda}^{\infty} \frac{dx}{1+x^2} \\ &= \frac{2}{\pi} \int_{\lambda}^{\infty} \frac{dx}{1+x^2}, \end{aligned}$$

where

$$\frac{2}{\pi} \int_{\lambda}^{\infty} \frac{dx}{2x^2} \leq \frac{2}{\pi} \int_{\lambda}^{\infty} \frac{dx}{1+x^2} \leq \frac{2}{\pi} \int_{\lambda}^{\infty} \frac{dx}{x^2}.$$

This implies that

$$\frac{2}{2\pi\lambda} \leq \mathbb{P}(|X| > \lambda) \leq \frac{2}{\pi\lambda},$$

which implies that $\lambda\mathbb{P}(|X| > \lambda)$ is constant. But $E[|X|] = +\infty!$

Theorem 4.43 (Kolmogorov's inequality). Let $(X_k)_{k \geq 1}$ be independent random variables with $E[X_k] = 0$ and $\text{Var}(X_k) < \infty$. Then for $\lambda > 0$, we have

$$\mathbb{P}\left(\max_{1 \leq k \leq n} |S_k| \geq \lambda\right) \leq \frac{\text{Var}(S_n)}{\lambda^2} = \frac{1}{\lambda^2} \sum_{k=1}^n E[X_k^2].$$

Example 4.44 (Fall 2018, #6). Let X_1, \dots, X_n, \dots be identically distributed (but, not necessarily independent) random variables with finite first moment. Is the following,

$$n^{-1} E\left[\max_{1 \leq k \leq n} |X_k|\right] \longrightarrow 0,$$

as $n \rightarrow \infty$, true or false?

(Uses dominated convergence.)

Example 4.45 (Fall 2016, #7). Let X be a random variable taking values on the interval $[1, 2]$. Find sharp lower and upper estimates on the quantity $E[X]E[\frac{1}{X}]$. Provide an example of a random variable for which each of the lower and upper estimates are obtained.

Hints:

- $ab \leq \frac{1}{2} \left(\frac{a}{2} + b\right)^2$;
- $ab \leq \frac{1}{4}(a+b)^2$.

4.6 Variance and covariance

Definition 4.46. We define the *variance* of a random variable X to be

$$\text{Var}(X) = E[X^2] - E[X]^2, \quad \sigma_X := \sqrt{\text{Var}(X)}.$$

For two random variables X, Y we define their covariance to be $\text{Cov}(X, Y) := E[(X - E[X])(Y - E[Y])]$ so that $\text{Cov}(X, X) = \text{Var}(X)$. If $X = (X_1, \dots, X_n)$ is a random vector, then we define its corresponding *covariance matrix* Cov_X to be the $n \times n$ symmetric matrix whose $(ij)^{\text{th}}$ entry is given by $\text{Cov}(X_i, X_j)$. We also define the *correlation* of X and Y to be $\text{Corr}(X, Y) = \text{Cov}(X, Y) / \sigma_X \sigma_Y$.

Proposition 4.47 (Properties). *We have the next properties of the variance and covariance:*

- (1) $\text{Var}(X) \geq 0$ with equality iff X is constant almost surely.
- (2) $\text{Var}(X + b) = \text{Var}(X), \forall b \in \mathbb{R}$.
- (3) For all $a \in \mathbb{R}$, $\text{Var}(aX) = a^2 \text{Var}(X)$.
- (4) $\text{Cov}(X + Z, Y) = \text{Cov}(X, Y) + \text{Cov}(Z, Y)$ and $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.
- (5) For any random variables X, Y , $-\sigma_X \sigma_Y \leq \text{Cov}(X, Y) \leq \sigma_X \sigma_Y$, where equality on the right is attained iff $Y = \alpha X + \beta$ for some $\alpha > 0$; and on the left iff $Y = -\alpha X + \beta$ for some $\alpha > 0$.
- (6) $\text{Corr}(X, Y) = \text{Corr}(aX + b, cY + d)$ for all $a, c \neq 0$.
- (7) If X, Y are integrable and independent, then $E[XY] = E[X]E[Y] \iff \text{Cov}(X, Y) = 0$.
- (8) If X_1, \dots, X_n are **independent** and in L^2 , then $\text{Var}(a_1 X_1 + \dots + a_n X_n) = a_1^2 \text{Var}(X_1) + \dots + a_n^2 \text{Var}(X_n)$.
- (9) $\text{Var}(\mathbb{1}_A) = \mathbb{P}(A)(1 - \mathbb{P}(A))$.
- (10) $\text{Cov}(\mathbb{1}_A, \mathbb{1}_B) = \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)$.

4.7 Uniform integrability

Definition 4.48. The sequence $(X_i)_{i \in I}$ is **uniformly integrable** if

$$\sup_{i \in I} E[|X_i|, |X_i| \geq R] \longrightarrow 0,$$

as $R \rightarrow \infty$. The interpretation of this definition is that the tails of the expectation are controlled uniformly.

Proposition 4.49. *If $(X_i)_{i \in I}$ are iid and X_1 is integrable, then (X_i) is uniformly integrable, i.e.,*

$$E[|X_i|, |X_i| \geq R] = E[|X_1|, |X_1| \geq R] \longrightarrow 0,$$

as $R \rightarrow \infty$.

Proof. For a fixed $R > 0$, let $Y_R := |X_1| \mathbf{1}_{|X_1| \geq R}$. Then by dominated convergence, $\forall R > 0$, $|Y_R| \leq |X_1| \implies Y_R \rightarrow 0$ as $R \rightarrow \infty$. This in turn implies that $E[Y_R] \rightarrow 0$ as $R \rightarrow \infty$. \square

Proposition 4.50. *If $\exists Y$ such that $|X_i| \leq Y$ for all $i \geq 1$ with Y integrable, then (X_i) is uniformly integrable.*

Proof. Since $\{|X_i| \geq R\} \subseteq \{Y \geq R\}$, we have the argument that

$$E[|X_i|, |X_i| \geq R] \leq E[Y, |X_i| \geq R] \leq E[Y, Y \geq R] \rightarrow 0,$$

as $R \rightarrow \infty$ (by integrability). So $\sup_{i \in I} E[|X_i|, |X_i| \geq R] \rightarrow 0$ as $R \rightarrow \infty$, as claimed. \square

Proposition 4.51. *If $\sup_{i \in I} E[|X_i|^\alpha] < \infty$ for some $\alpha > 1$, then $(X_i)_{i \in I}$ is uniformly integrable.*

Proof. We see that

$$\begin{aligned} E[|X_i|, |X_i| \geq R] &= E[|X_i| \mathcal{X}_{|X_i| \geq R}] \leq E \left[|X_i| \left(\frac{|X_i|}{R} \right)^{\alpha-1} \mathcal{X}_{|X_i| \geq R} \right] \\ &= E \left[\frac{|X_i|^\alpha}{R^{\alpha-1}} \mathcal{X}_{|X_i| \geq R} \right] \leq \frac{E[|X_i|^\alpha]}{R^{\alpha-1}}, \end{aligned}$$

which implies that

$$\sup_{i \in I} E[|X_i|, |X_i| \geq R] \leq \sup_{i \in I} \frac{E[|X_i|^\alpha]}{R^{\alpha-1}} \rightarrow 0,$$

as $R \rightarrow \infty$. \square

Example 4.52 (Spring 2019, #4). Let $(X_n)_{n \geq 1}$ be a sequence of non-negative uniformly integrable random variables such that, as $n \rightarrow +\infty$, $X_n \implies X$. Show that X is integrable and that $\lim_{n \rightarrow +\infty} E[X_n] = E[X]$.

(**Note:** The given proof is *very* careful with respect to tying up loose ends with analysis. Uses Fatou's lemma.)

(**Fact:** By uniform integrability, $(EX_n)_{n \geq 1}$ is bounded!)

4.8 Convergence and expectation

First, some preliminary notes: we recall the subset containment properties of limsups of sets given in Proposition 1.5. Also, we restate the precise definition of an *exponential random variable* here for clarity of exposition as: $X \sim \text{Exp}(\lambda)$ is exponential if μ_X has a density function of the form

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0; \\ 0, & x < 0. \end{cases}$$

Now we arrive at the focus of this subsection which is to dig into the next problem on iid exponential random variables:

Problem Statement: We claim that if X_1, \dots, X_n, \dots are iid with $X_1 \sim \text{Exp}(1)$, then

$$\mathbb{P} \left(\limsup_{n \rightarrow \infty} \frac{X_n}{\log n} = 1 \right) = 1$$

We notice that a proof of this result is equivalent to showing that

$$\text{(A)} \quad \mathbb{P} \left[\limsup_{n \rightarrow \infty} \frac{X_n}{\log n} > 1 + \delta \right] = 0, \forall \delta > 0; \text{ and then similarly that}$$

$$\text{(B)} \quad \mathbb{P} \left[\limsup_{n \rightarrow \infty} \frac{X_n}{\log n} > 1 - \delta \right] = 1, \forall \delta > 0.$$

Note next that

$$\left\{ \limsup_{n \rightarrow \infty} \frac{X_n}{\log n} = 1 \right\} = \bigcap_{\delta > 0} \left\{ 1 + \delta > \limsup_{n \rightarrow \infty} \frac{X_n}{\log n} > 1 - \delta \right\} = \bigcap_{k \geq 1} A_k,$$

where

$$A_k := \left\{ 1 + \frac{1}{k} > \limsup_{n \rightarrow \infty} \frac{X_n}{\log n} > 1 - \frac{1}{k} \right\},$$

for $k \geq 1$. Also, we have that $A_k \searrow \left\{ \limsup_{n \rightarrow \infty} \frac{X_n}{\log n} = 1 \right\}$. Moving along, we can see that

$$\begin{aligned} \mathbb{P} \left[\frac{X_n}{\log n} > 1 + \delta - \varepsilon \right] &= e^{-(1+\delta-\varepsilon)\log n} = \frac{1}{n^{1+\delta-\varepsilon}} \\ \implies \sum_{n \geq 1} \mathbb{P} \left[\frac{X_n}{\log n} > 1 + \delta - \varepsilon \right] &= \sum_{n \geq 1} \frac{1}{n^{1+\delta-\varepsilon}} < \infty, \quad \text{if } \varepsilon < \delta. \end{aligned}$$

So in summary, to complete the proof of **Q**, we use Borel-Cantelli in the form of

$$\mathbb{P} \left[\left\{ \limsup_{n \rightarrow \infty} \frac{X_n}{\log n} > 1 + \delta \right\} \right] \subseteq \mathbb{P} \left[\limsup_{n \rightarrow \infty} \left\{ \frac{X_n}{\log n} > 1 + \delta \right\} \right] = 0.$$

To show **P**, we need to consider lower bounds: $\forall \varepsilon > 0$,

$$\limsup_{n \rightarrow \infty} \left\{ \frac{X_n}{\log n} > 1 - \delta + \varepsilon \right\} \subseteq \left\{ \limsup_{n \rightarrow \infty} \frac{X_n}{\log n} > 1 + \delta \right\}.$$

To use the second part of Borel-Cantelli on independent random variables, we want that

$$\sum_{n \geq 1} \mathbb{P} \left[\frac{X_n}{n} > 1 - \delta + \varepsilon \right] = \sum_{n \geq 1} \frac{1}{n^{1-\delta+\varepsilon}} = +\infty,$$

which happens if $\varepsilon \leq \delta$. Then this implies that

$$\mathbb{P} \left[\limsup_{n \rightarrow \infty} \left\{ \frac{X_n}{\log n} > 1 + \delta - \varepsilon \right\} \right] = 1, \quad \text{for } \varepsilon \leq \delta \implies \mathbf{P}.$$

Example 4.53 (Spring 2018, #1). Let $(X_n)_{n \geq 1}$ be a sequence of iid random variables with $X_1 \sim \text{Exp}(\lambda)$, i.e., for $t \geq 0$ we have that $\mathbb{P}(X_n \geq t) = e^{-\lambda t}$ where $\lambda > 0$. Prove that

$$\limsup_{n \rightarrow \infty} \frac{X_n}{\log n} < \infty, \quad \text{a.s.}$$

(**Hint:** Think easy application of Borel-Cantelli.)

4.9 Characteristic functions of a random variable

Definition 4.54 (Characteristic Function). The *characteristic function* of a random variable X is the function $\phi_X : \mathbb{R} \rightarrow \mathbb{C}$ defined by $\phi_X(t) := E[e^{iXt}]$.

We have the following properties of the characteristic function ϕ_X of any random variable X :

- We have that $\phi_X(0) = 1$ and for all t : $|\phi_X(t)| = |E[e^{iXt}]| \leq E[|e^{iXt}|] = 1$;

- The characteristic function $\phi_X(t)$ is a uniformly continuous function of t (for any $t \in \mathbb{R}$).
- If $Y = aX + b$ for constants $a, b \in \mathbb{R}$, then

$$\phi_Y(t) = E[e^{i(aX+b)t}] = e^{ibt} \phi_X(at).$$

- The function $M_X(t) = e^{tX}$ is called the **moment generating function** of X . Unlike the characteristic function, *this GF does not necessarily exist* for all $t \in \mathbb{R}$. For those t such that $M_X(t)$ does converge, we have the relation that $\phi_X(t) = M_X(it)$;
- We have that for $k > 0$ (not necessarily an integer),

$$\mathbb{E}[X^k] \leq \frac{\Gamma(k+1)}{t^k} M_X(t).$$

- If X has a density f , then

$$\phi_X(t) = \int_{-\infty}^{\infty} e^{itx} f(x) dx,$$

which is the *Fourier transform* of f at $-t$;

- If two random variables have the same characteristic function, then they have the same underlying distribution;
- If X_1, \dots, X_n are independent random variables, then

$$\phi_{X_1+\dots+X_n}(t) = \phi_{X_1}(t) \cdots \phi_{X_n}(t).$$

- If $X \sim N(0, 1)$, then $\phi_X(t) = e^{-t^2/2}$. Similarly, if $Y \sim N(\mu, \sigma^2)$, then Y has the same distribution as $\sigma X + \mu$, and hence, $\phi_Y(t) = e^{i\mu t} e^{-(\sigma t)^2/2}$.
- If the characteristic functions of X_n converge pointwise to that of X , then $X_n \implies X$ (weakly).
- If X, Y are independent, then the characteristic function of $X + Y$ equals the product of their characteristic functions: $\phi(u) = \phi_X(u)\phi_Y(u)$.
- If X_1, \dots, X_n are iid and $\bar{S}_n := \frac{X_1+\dots+X_n}{n}$, then $\phi_{\bar{S}_n}(t) = \phi_{X_1}(\frac{t}{n})^n$.

Distribution	Ch. Func	MGF
Ber(p)	$1 - p + pe^{it}$	$1 - p + pe^t$
Binom(n, p)	$(1 - p + pe^{it})^n$	$(1 - p + pe^t)^n$
Poisson(λ)	$\exp(\lambda(e^{it} - 1))$	$\exp(\lambda(e^t - 1))$
Uniform(a, b)	$\frac{1}{it(b-a)} [e^{ibt} - e^{iat}]$	$\frac{1}{t(b-a)} [e^{bt} - e^{at}]$
Exp(λ)	$(1 - \frac{it}{\lambda})^{-1}$	$(1 - \frac{t}{\lambda})^{-1}$
$N(0, 1)$	$e^{-t^2/2}$	$e^{t^2/2}$
$N(0, \sigma^2)$	$e^{-(\sigma t)^2/2}$	$e^{(\sigma t)^2/2}$
$N(\mu, \sigma^2)$	$e^{i\mu t - (\sigma t)^2/2}$	$e^{\mu t + (\sigma t)^2/2}$

Proposition 4.55 (Moments of a Random Variable). *Let ϕ be the characteristic function of X . If $E[|X|] < \infty$, then ϕ is continuously differentiable, and*

$$\phi'(0) = iE[X].$$

If $E[|X|^k] < \infty$ for some positive integer $k \geq 1$, then ϕ has k continuous derivatives and

$$\phi^{(j)}(0) = i^j E[X^j], \quad \forall 0 \leq j \leq k.$$

Theorem 4.56 (Inversion). *The PDF of X , $f_X(x)$, is given by the inversion formula*

$$f_X(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \phi_X(t) dt.$$

Similarly, if g is a Borel function such that $g(X)$ is integrable, then

$$\mathbb{E}[g(X)] = \frac{1}{2\pi} \int_{\mathbb{R}} G(t) \phi_X(t) dt,$$

where

$$G(t) = \int_{\mathbb{R}} g(x) e^{-itx} dx.$$

We also have that

$$\mu\{x : |x| > 2/u\} \leq \frac{1}{u} \int_{-u}^u (1 - \phi(t)) dt.$$

- If $\mathbb{E}X^2 < \infty$, then as $t \rightarrow 0$

$$\phi(t) = 1 + it\mathbb{E}X - \frac{t^2}{2}\mathbb{E}X^2 + o(t^2).$$

- If $\limsup_{h \searrow 0} \frac{1}{h^2} [\phi(h) - 2\phi(0) + \phi(-h)] > -\infty$, then $\mathbb{E}|X|^2 < \infty$.

Examples involving characteristic functions

Example 4.57 (Spring 2018, #9). Find an example of a random variable with a density function, but whose characteristic function ϕ_X satisfies

$$\int_{-\infty}^{\infty} |\phi_X(t)| dt = \infty.$$

Example 4.58 (Fall 2018, #4). Let ϕ be the characteristic function of a random variable X . Show that

$$\Psi_1(t) = |\phi(t)|^2, \Psi_2(t) = \frac{1}{t} \int_0^t \phi(s) ds,$$

are also characteristic functions.

(**Hint:** $\int_0^1 \mathbb{E}(e^{itXU} | U = u) du = \mathbb{E}\mathbb{E}(e^{itXU} | U)$.)

Example 4.59 (Spring 2017, #1). Show that if X_n and Y_n are independent for $n = 1, 2, \dots$ and $X_n \implies X$ and $Y_n \implies Y$, where X, Y are independent, then $X_n + Y_n \implies X + Y$.

Example 4.60 (Spring 2017, #5). Show that the characteristic function ϕ of a random variable is real if and only if X and $-X$ have the same distribution.

Example 4.61 (Fall 2016, #5). Let ξ_1, ξ_2 be independent random variables with respective characteristic functions

$$\phi_1(u) = \frac{1 - iu}{1 + u^2}, \phi_2(u) = \frac{1 + iu}{1 + u^2}.$$

Find the probability that $\xi_1 + \xi_2$ takes values in $(3, \infty)$.

Example 4.62 (Spring 2016, #5). Let $(N_t)_{t \geq 0}$ be a rate- λ Poisson process. Let X_1, X_2, \dots be iid random variables with $E[X_1] < \infty$, and define

$$S_t := \sum_{i=1}^{N_t} X_i.$$

Show that S_t/t converges in probability to a constant and compute this constant.

4.10 The Levy continuity theorem

We assume the following hypotheses:

- For $(X_n)_{n \geq 1}$ a sequence of random variables, we define the respective characteristic functions by $\phi_n(t) := \mathbb{E}[e^{itX_n}]$;
- Suppose that $\phi_n(t) \rightarrow \phi(t)$ pointwise for all $t \in \mathbb{R}$.

Then the *Levy continuity theorem* states that TFAE:

- $X_n \Longrightarrow X$, i.e., the cdfs converge at every point of continuity;
- The sequence $(X_n)_{n \geq 1}$ is so-called *tight*:

$$\lim_{x \rightarrow \infty} \left(\sup_n \mathbb{P}[|X_n| > x] \right) = 0;$$

- The function $\phi(t)$ is the characteristic function of some (limiting) random variable X ;
- $\phi(t)$ is continuous in t ;
- $\phi(t)$ is continuous at $t := 0$.

5 Modes of convergence

5.1 Definitions of convergence

Definition 5.1 (Convergence Everywhere). Let $X_n, X : \Omega \rightarrow \mathbb{R}$. Then $X_n \rightarrow X$ if $X_n(\omega) \rightarrow X(\omega)$ as $n \rightarrow \infty$ for all $\omega \in \Omega$.

Definition 5.2 (Almost Sure Convergence). We say that an event $A \subseteq \Omega$ is *almost sure* if $\mathbb{P}(A) = 1$, while noting that this property is distinguished from “provably certain” in so much as there may be other outcomes in the sample space which lie outside of A , but these events are in zero measure subsets of the field. Now, we write $X_n \xrightarrow{a.s.} X$ if $\mathbb{P}(\{\omega : X_n(\omega) \rightarrow X(\omega)\}) = 1$. Equivalently, we define that $X_n \xrightarrow{a.s.} X$ if $\exists A \in \mathcal{F}$ an almost sure event such that $X_n(\omega) \rightarrow X(\omega), \forall \omega \in A$. We have almost sure convergence if

$$\limsup_{n \rightarrow \infty} X_n = \liminf_{n \rightarrow \infty} X_n,$$

in which case we may write that

$$X = \lim_{n \rightarrow \infty} X_n = \limsup_{n \rightarrow \infty} X_n = \liminf_{n \rightarrow \infty} X_n.$$

Proposition 5.3. $X_n \xrightarrow{a.s.} X \iff \forall \varepsilon > 0, \mathbb{P}(\limsup_{n \rightarrow \infty} \{|X_n - X| > \varepsilon\}) = 0$. An important corollary is that if $\forall \varepsilon > 0 \sum_{n \geq 1} \mathbb{P}(|X_n - X| > \varepsilon) < \infty$, then $X_n \xrightarrow{a.s.} X$. (**)

Definition 5.4 (Convergence in Probability). We write $X_n \xrightarrow{\mathbb{P}} X$ if $\forall \varepsilon > 0, \mathbb{P}(|X_n - X| > \varepsilon) \rightarrow 0$. Note that convergence in probability implies finiteness (a.e.) almost surely.

Definition 5.5 (Convergence in Mean, a.k.a., L^p Convergence). We write $X_n \xrightarrow{L^p} X$ if $E[|X_n - X|^p] \rightarrow 0$ as $n \rightarrow \infty$.

Definition 5.6 (Weak Convergence, a.k.a., Convergence in Distribution). A sequence of random variables $(X_n)_{n \geq 1}$ converges in distribution, or converges weakly, to X if

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

for every $x \in \mathbb{R}$ at which F_X is continuous. In this case, we write $X_n \xrightarrow{D} X$, or sometimes also $X_n \xrightarrow{\text{dist}} X$.

Proposition 5.7 (Equivalent Conditions for Convergence in Distribution). For any sequence $(X_n)_{n \geq 1}$ of random variables, we have that $X_n \xrightarrow{\text{dist}} X$ if and only iff any of the following conditions are met:

- (a) $E[f(X_n)] \rightarrow E[f(X)]$, for all bounded, continuous functions f ;
- (b) $E[f(X_n)] \rightarrow E[f(X)]$, for all bounded, Lipschitz functions f . Here, we say that a function $f : \mathbb{R} \rightarrow \mathbb{R}$ is Lipschitz, or 1-Lipschitz, if there exists a constant $K > 0$ which is independent of $x, y \in \mathbb{R}$ such that $\forall x, y: |f(x) - f(y)| \leq K \cdot |x - y|$;
- (c) $\limsup_{n \rightarrow \infty} \{\mathbb{P}(X_n \in C)\} \leq \mathbb{P}(X \in C)$, for all closed sets $C \subseteq \Omega$.

Example 5.8 (Fall 2018, #7). Let X_1, X_2, \dots be iid random variables with common characteristic function ϕ , and let $S_n = \sum_{k=1}^n X_k$. Show that if ϕ is differentiable at 0 with $\phi'(0) = i\mu$, then, as $n \rightarrow \infty, S_n/n \rightarrow \mu$ in probability.

Hints: Proof uses Lévy continuity theorem. Always we have $\phi(0) = 1$, and here we have that $\phi'(0) = \lim_{n \rightarrow \infty} \frac{\phi(t/n) - 1}{t/n - 0}$.

Example 5.9 (Fall 2016, #8). Show that for a sequence of random variables X_n , one has that $X_n \xrightarrow{\mathbb{P}} X$ in probability if and only if

$$E \left[e^{\min(2, |X_n - X|)} - 1 \right] \longrightarrow 0,$$

as $n \rightarrow \infty$.

Hints: 1) See the first argument; 2) The reverse direction uses Markov.

Example 5.10 (Spring 2016, #4). Let X_1, X_2, \dots be iid exponential variables with parameter $\lambda := 1$, and set

$$M_n := \max(X_1, \dots, X_n).$$

Find sequences $(a_n), (b_n) \subset \mathbb{R}$ such that $(M_n - a_n)/b_n$ converges in distribution.

Example 5.11 (Spring 2015, #6). Assume $(X_n)_{n \geq 1}$ is a sequence of iid random variables with mean 0 and variance 1. Show that

$$Y_n = \frac{\sqrt{n}X_1 + \sqrt{n-1}X_2 + \dots + X_n}{n},$$

converges weakly (in distribution) to a normal $N(0, 1/2)$.

Example 5.12 (Spring 2015, #7). Assume that $(U_n)_{n \geq 1}$ is a sequence of uniform random variables on $[0, 1]$. Let $V_n := \max(U_1, U_2^2, \dots, U_n^n)$. Show that $(1 - V_n) \log(n)$ converges weakly (in distribution) to an exponential random variable with parameter $\lambda = 1$.

5.2 Remarks on the equivalence of studying convergence to zero

Proposition 5.13 (Convergence to Zero). *If $(Y_n)_{n \geq 1}$ is any sequence of random variables, then $Y_n \xrightarrow{a.s.} 0 \iff \forall \varepsilon > 0, \mathbb{P}(\limsup |Y_n| > \varepsilon) = 0$. And then as a consequence, $X_n \xrightarrow{a.s.} X \iff \forall \varepsilon > 0, \mathbb{P}(\limsup |X_n - X| > \varepsilon) = 0$. (**)*

Remark 5.14 (Notes on Convergence to Zero). We notice that it is sufficient to study convergence of a random variable to zero since we can always replace the expression $X_n \xrightarrow{a.s.} X$ by $Y_n := X_n - X \xrightarrow{a.s.} 0$. Next, we have the observation that for any sequence $(a_n)_{n \geq 1}$,

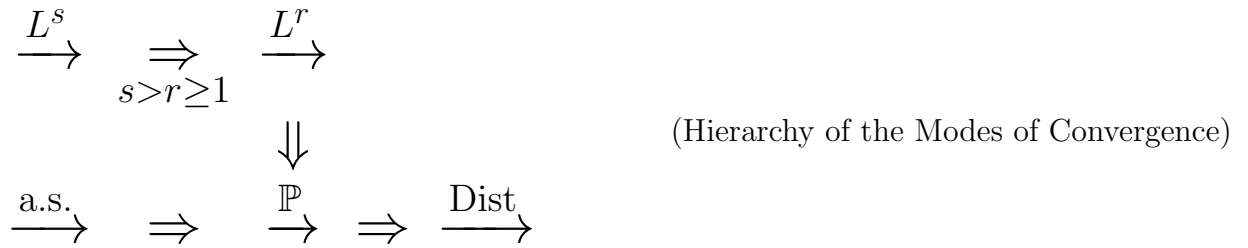
$$\begin{aligned} a_n \longrightarrow 0 &\iff \limsup_{n \rightarrow \infty} |a_n| = 0 \\ &\iff \inf_{n \geq 1} \sup_{k \geq n} a_k = 0 \\ &\iff \forall \varepsilon > 0, \exists N_\varepsilon : \sup_{k \geq N_\varepsilon} a_k < \varepsilon. \end{aligned}$$

Remark 5.15 (Borel-Cantelli as a Practical Tool for Proving Convergence). As a tool for proving almost sure convergence, we can apply Borel-Cantelli: If $\forall \varepsilon > 0, \sum_{n \geq 1} \mathbb{P}(|X_n - X| > \varepsilon) < \infty$, then $X_n \xrightarrow{a.s.} X$. Note that the convergence of the above sum, at minimum requires that $\mathbb{P}(|X_n - X| > \varepsilon) \rightarrow 0$ as $n \rightarrow \infty$ for all $\varepsilon > 0$. We also point out the following key observations:

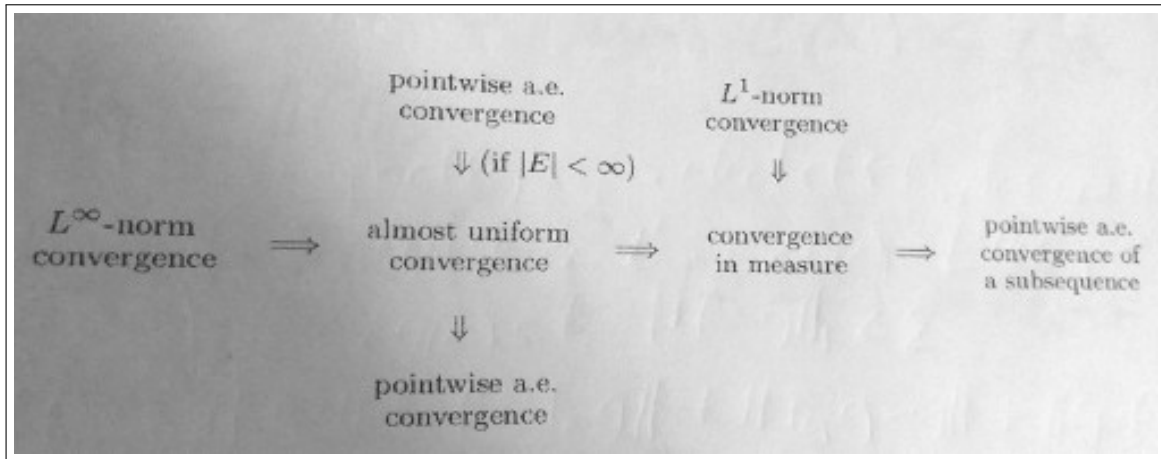
- If $\forall \varepsilon > 0: \sum_{n \geq 1} \mathbb{P}(|X_n - X| > \varepsilon) < \infty$, then $\mathbb{P}(\limsup\{|X_n - X| > \varepsilon\}) = 0$, for all $\varepsilon > 0$;
- If $\sum_{n \geq 1} \mathbb{P}(|X_n - X| > \varepsilon) < \infty, \forall \varepsilon > 0$, then $X_n \xrightarrow{a.s.} X$, where we consider $\mathbb{P}(|X_n - X| > \varepsilon)$ to be the *tail probability* which we must estimate for n large.

5.3 Flowcharts of implications: convergence hierachy

We have the following map of convergence implications:



We compare this to the real analysis modes summary chart reproduced as follows from Heil's notes:



Proposition 5.16 (Almost Everywhere Convergence of a Subsequence). *Convergence in probability implies that there exists a subsequence (k_n) of the original sequence which almost surely converges:*

$$\exists \{k_n\}_{n \geq 1} \subseteq \mathbb{N} : X_n \xrightarrow{\mathbb{P}} X \implies X_{k_n} \xrightarrow{\text{a.s.}} X.$$

Disproof: Counterexample to Convergence in Probability Implies Almost Sure Convergence. We will demonstrate a sequence of random variables that converges in probability to zero on $[0, 1]$, but which does not converge almost surely to zero. For $n \geq 1$ and $1 \leq j \leq n$, define

$$S_{n,j} := \left[\frac{j-1}{n}, \frac{j}{n} \right] \subseteq [0, 1],$$

and set $X_n(x) := \chi_{S_{n,j}}(x)$. Let $\varepsilon > 0$ and observe that

$$\mathbb{P}(\{x \in [0, 1] : f_n(x) > \varepsilon\}) = \mathbb{P}(\{x \in [0, 1] : f_n(x) = 1\}) = |S_{n,j}| = \frac{1}{n} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

On the other hand, given any $x \in [0, 1]$ there are infinitely-many $n \in \mathbb{N}$ such that $x \in S_{n,j}$, i.e., such that $X_n(x) = 1$. This implies that there is a subsequence $\{X_{n_k}\}$ such that $X_{n_k} \xrightarrow{\mathbb{P}} 1$ and hence $X_n \not\rightarrow 0$ almost surely on $[0, 1]$. \square

Proof: Existence of an Almost Surely Convergent Subsequence. Suppose that $X_n \xrightarrow{\mathbb{P}} X$. We need to find a subsequence $\{X_{n_k}\}_{k \geq 1}$ such that $X_{n_k} \xrightarrow{\text{a.s.}} X$ as $k \rightarrow \infty$ for all $\omega \in A$ where $A \subseteq \Omega$ is some large enough event in the sample space such that $\mathbb{P}(A) = 1$. For $j \geq 1$, choose L_j such that for all $k \geq L_j$

$$\mathbb{P}(\{\omega : |X_k - X|(x) \geq 1/j\}) < \frac{1}{2^j}.$$

We can just as well assume here that the L_j 's satisfy: $L_1 < L_2 < L_3 < \dots$. For $j \geq 1$, we define

$$E_j := \{\omega : |X_{L_j} - X|(\omega) \geq 1/j\},$$

and set

$$Z := \limsup_{j \rightarrow \infty} E_j = \bigcap_{m \geq 1} \bigcup_{j \geq m} E_j.$$

Now for $m \geq 1$, we can see that

$$\mathbb{P}(Z) \leq \mathbb{P}\left(\bigcup_{j \geq m} E_j\right) \leq \sum_{j \geq m} \mathbb{P}(E_j) < \sum_{j \geq m} 2^{-j} = 2^{1-m} \rightarrow 0,$$

as $m \rightarrow \infty$. So we conclude that $\mathbb{P}(Z) = 0$. Next, if $\omega \in \Omega \setminus Z$, then $\omega \notin \bigcup_{j \geq m} E_j$ for some $m \geq 1$. This implies that $\omega \notin E_j$ for all $j \geq m$. Thus $|X_{L_j} - X|(\omega) \leq 1/j$ for $j \geq m$, which implies that $\lim_{j \rightarrow \infty} X_{L_j}(\omega) = X(\omega)$ for all $\omega \in \Omega \setminus Z$. So it suffices to take $X_{n_j} := X_{L_j}$ so that $X_{n_k} \xrightarrow{a.s.} X$, or equivalently with $A := \Omega \setminus Z$ so that $X_{n_k}(\omega) \rightarrow X(\omega)$ for all $\omega \in A$ where, as proved above, $\mathbb{P}(A) = \mathbb{P}(\Omega) - \mathbb{P}(Z) = 1 - 0 = 1$. \square

5.4 Convergence theorems (analogs to real analysis)

Theorem 5.17 (Lebesgue's BCT). *Let f_n be integrable $\forall n \geq 1$ such that (a) $f_n \xrightarrow{\mathbb{P}} f$; or (b) $f_n \xrightarrow{a.s.} f$ for some measurable function f . If $|f_n(x)| \leq g(x)$ a.e. for all n where g is integrable, then f is integrable and*

$$\lim_{n \rightarrow \infty} \int |f_n - f| d\mu = 0.$$

Theorem 5.18. *If f, g are measurable such that $|f| \leq g$ and g is integrable, then f is integrable.*

Theorem 5.19 (Monotone convergence theorem). *Let $\{f_k\}_{k \geq 1}$ be a sequence of measurable functions on E . Then*

(1) *If $f_k \nearrow f$ a.s. on E and $\exists \varphi$ integrable such that $f_k \geq \varphi$ a.e. in E for all $k \geq 1$, then*

$$\lim_{k \rightarrow \infty} \int_E f_k = \int_E f.$$

(2) *If $f_k \searrow f$ a.s. on E and $\exists \varphi$ integrable such that $f_k \leq \varphi \forall k \geq 1$, then*

$$\lim_{k \rightarrow \infty} \int_E f_k = \int_E f.$$

Theorem 5.20 (General form of LDCT). *Let f_k be measurable for all $k \geq 1$. Suppose that (a) $\lim_{k \rightarrow \infty} f_k = f$ a.e. in E ; and (b) $\exists \varphi$ integrable such that for all $k \geq 1$, $|f_k| \leq \varphi$ a.e. in E . Then*

$$\lim_{k \rightarrow \infty} \int f_k = \int f.$$

Lemma 5.21 (Fatou). *Suppose that $E \subseteq \mathbb{R}^n$ is measurable and let $f_k \geq 0$ be measurable for all $k \geq 1$. Then*

$$\liminf_{k \rightarrow \infty} \int_E f_k \geq \int_E \left(\liminf_{k \rightarrow \infty} f_k \right).$$

We can obtain the same conclusion if we instead assume that the $f_k \geq \varphi$ for all $k \geq 1$ when φ is integrable on E . Notice that the statement of Fatou's lemma makes no a priori assumptions on the convergence of the sequence $\{f_k\}_{k \geq 1}$.

Theorem 5.22 (Uniform convergence theorems). *We have the following variants of uniform convergence theorems:*

- $f_n \rightarrow f$ uniformly on $[a, b]$ with f_n all continuous $\implies f$ is continuous.
- If f_n is differentiable on $[a, b]$ and $\lim_{n \rightarrow \infty} f_n(x_0)$ exists for some $x_0 \in [a, b]$ and f'_n converge uniformly on $[a, b]$, then $f_n \rightarrow f$ uniformly on $[a, b]$ and $f'(x) = \lim_{n \rightarrow \infty} f'_n(x)$ for all $x \in [a, b]$.
- If f_n is integrable on $[a, b]$ and $f_n \rightarrow f$ uniformly, then f is integrable and

$$\int_{[a,b]} f = \lim_{n \rightarrow \infty} \int_{[a,b]} f_n.$$

5.5 Properties of the modes of convergence for random variables

Fact: If $\mu(X) < \infty$, then $L^q(\mu) \subsetneq L^p(\mu)$ for any $0 < p < q \leq \infty$. This can be proved by applying Hölder with the conjugate exponents $\boxed{p_0 := q/p \text{ and } q_0 := q/(q-p)}$.

Proposition 5.23 (Cauchy-Schwarz and Hölder inequalities). *Note that we assume that the functions f, g are both square integrable, i.e., $|f|^2, |g|^2$ are both integrable. If this is the case then,*

$$\begin{aligned} \int_0^1 |f| &\leq \left(\int_0^1 |f|^2 \right)^{1/2} \\ \left| \int fg \right| &\leq \left(\int |f|^2 \right)^{1/2} \left(\int |g|^2 \right)^{1/2}. \end{aligned}$$

Note that the latter equation above is the special case of the more general cases in Hölder's inequality for $\frac{1}{p} + \frac{1}{q} = 1 \iff p + q = pq$ when $p = q = 2$:

$$\left| \int fg \right| \leq \|f\|_p \cdot \|g\|_q.$$

Dual (conjugate) exponents: $\boxed{\text{If } p + q = pq, \text{ then } p = q/(q-1)}$.

Note that *Minkowski's theorem* states that $\|f + g\|_p \leq \|f\|_p + \|g\|_p$. Minkowski also implies that

$$\left\| \int f \right\|_p \leq \int \|f\|_p.$$

Proof of Minkowski. We let p be arbitrary with $f, g \in L^p$ and observe that:

$$\begin{aligned} \|f + g\|_p^p &= \int |f + g|^p = \int |f + g|^{p-1} |f + g| \\ &\leq \int |f + g|^{p-1} (|f| + |g|) d\mu, \quad \text{then we apply Hölder} \\ &\leq (\|f\|_p + \|g\|_p) \left(\int |f + g|^{(p-1)\frac{p}{p-1}} \right)^{1-1/p} \\ &= (\|f\|_p + \|g\|_p) \frac{\|f + g\|_p^p}{\|f + g\|_p}. \end{aligned}$$

□

If $\alpha > 0$, we can obtain that

$$\mu(\{x : |f(x)| > \alpha\}) \leq \left(\frac{\|f\|_p}{\alpha}\right)^p.$$

Another important result which is non-trivial to prove and due to Riesz is that for $1 \leq q < \infty$ and $g \in L^1$:

$$\|g\|_q = \sup \left\{ \left| \int fg \right| : \|f\|_p = 1 \right\}, \quad \text{where } \frac{1}{p} + \frac{1}{q} = 1.$$

Example 5.24 (Dense Class Arguments). *Compactly supported continuous functions are dense in L^1 : $C_c[a, b] = L^1[a, b]$. Other dense functions in L^p are the simple (i.e., staircase) functions, and polynomials. If $f \in L^p$ and C is dense in L^p then $\forall \varepsilon > 0, \exists g \in C$ such that $\|f - g\|_p < \varepsilon$.*

Proposition 5.25. *Provided that the probability space is complete:*

- If $X_n \xrightarrow{P} X$ and $X_n \xrightarrow{P} Y$, then $X = Y$ almost surely.
- If $X_n \xrightarrow{a.s.} X$ and $X_n \xrightarrow{a.s.} Y$, then $X = Y$ almost surely.
- If $X_n \xrightarrow{L^r} X$ and $X_n \xrightarrow{L^r} Y$, then $X = Y$ almost surely.
- If $X_n \xrightarrow{P} X, Y_n \xrightarrow{P} Y$, then for any $a, b \in \mathbb{R}$, $aX_n + bY_n \xrightarrow{P} aX + bY$, and $X_n Y_n \xrightarrow{P} XY$.
- If $X_n \xrightarrow{a.s.} X, Y_n \xrightarrow{a.s.} Y$, then for any $a, b \in \mathbb{R}$, $aX_n + bY_n \xrightarrow{a.s.} aX + bY$, and $X_n Y_n \xrightarrow{a.s.} XY$.
- If $X_n \xrightarrow{L^r} X, Y_n \xrightarrow{L^r} Y$, then for any $a, b \in \mathbb{R}$, $aX_n + bY_n \xrightarrow{L^r} aX + bY$.
- None of the above statements are true for weak convergence in distribution.
- Convergence in distribution to a constant implies convergence in probability.

6 Laws of large numbers

6.1 The weak law of large numbers

Example 6.1 (Key Observations). Suppose that $(X_i)_{i \geq 1}$ have the same L^2 norm and that $E[X_i] = 0$ for all $i \geq 1$. Then:

- $\text{Var}\left(\frac{X_1 + \dots + X_n}{n}\right) = E\left[\left|\frac{X_1 + \dots + X_n}{n}\right|^2\right]$, which implies that

$$\text{Var}\left(\frac{X_1 + \dots + X_n}{n}\right) = \left\| \frac{X_1 + \dots + X_n}{n} \right\|_{L^2}^2 \leq \frac{1}{n} (\|X_1\|^2 + \dots + \|X_n\|^2) = \|X_1\|_{L^2}^2.$$

- If $X_1 = \dots = X_n$, then $\text{Var}\left(\frac{X_1 + \dots + X_n}{n}\right) = \text{Var}(X_1)$.
- If X_1, \dots, X_n are iid, then

$$\text{Var}\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{1}{n^2} \sum_{i,j=1}^n \text{Cov}(X_i, X_j) = \frac{1}{n^2} \sum_{m=1}^n \text{Var}(X_m) = \frac{1}{n} \text{Var}(X_1).$$

Theorem 6.2 (WLLN). If X_1, \dots, X_n are iid with mean μ and variance σ^2 , then $\bar{S}_n := \frac{X_1 + \dots + X_n}{n} \xrightarrow{P, L^2} \mu$ as $n \rightarrow \infty$.

Proof. First, we show the statement of L^2 convergence. Since $\mu = E[\bar{S}_n]$:

$$E[(\bar{S}_n - \mu)^2] = \text{Var}(\bar{S}_n) = \frac{\sigma^2}{n} \rightarrow 0,$$

as $n \rightarrow \infty$. So $\bar{S}_n \xrightarrow{L^2} \mu$. Next, for all $\varepsilon > 0$,

$$\mathbb{P}(|\bar{S}_n - \mu| > \varepsilon) \leq \frac{E[(\bar{S}_n - \mu)^2]}{\varepsilon^2} \leq \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0,$$

as $n \rightarrow \infty$. Hence, $\bar{S}_n \xrightarrow{P} \mu$. □

Theorem 6.3 (WLLN, V2). Suppose that X_1, \dots, X_n are iid with mean μ and $E[X_1^4] < \infty$. Then $\bar{S}_n \xrightarrow{L^4, a.s.} \mu$.

Proof. Since we can replace X_i by $Y_i := X_i - \mu$ and bound the resulting norm of this random variable, we can assume that $\mu = 0$. Using Markov,

$$\mathbb{P}(|\bar{S}_n| > \varepsilon) = \mathbb{P}(\bar{S}_n^4 > \varepsilon^4) \leq \frac{E[\bar{S}_n^4]}{\varepsilon^4} \leq \frac{C}{n^2\varepsilon^4}.$$

Thus $\bar{S}_n \xrightarrow{P} 0$. And by Borel-Cantelli, we have that

$$\sum_{n \geq 1} \mathbb{P}(|\bar{S}_n| \geq \varepsilon) \leq \sum_{n \geq 1} \frac{C}{n^2\varepsilon^4} < \infty,$$

and so $\bar{S}_n \xrightarrow{a.s.} 0$. □

Theorem 6.4. If X_1, \dots, X_n are independent with mean μ and are uniformly integrable, then $\bar{S}_n \xrightarrow{L^1, P} \mu$.

Example 6.5 (Spring 2019, #3). Assume that $\{X_n\}_{n \geq 1}$ are random variables such that

1. $E[X_n] = 0$ and $E[X_n^2] \leq 1$ for any $n \geq 1$;
2. $E[X_i X_j] \leq 0$ for any $i \neq j$.

Show that for any sequence $\{a_n\}_{n \geq 1} \subset [1/2, 2]$,

$$\frac{a_1 X_1 + a_2 X_2 + \dots + a_n X_n}{a_1 + a_2 + \dots + a_n} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

(*Hints: Use Chebyshev.*)

6.2 Large deviations

If X_1, \dots, X_n are iid with mean 0 and $E[e^{\alpha|X_1|}] < \infty$ for some $\alpha > 0$, then

$$\mathbb{P}(|\bar{S}_n| > \varepsilon) \leq 2e^{-nI(\varepsilon)},$$

where the **rate function** $I(\varepsilon)$ is given by

$$I(\varepsilon) = \sup_{\lambda \in \mathbb{R}} \{\varepsilon \lambda - \log E[e^{\lambda X_1}]\}.$$

The function $f(\lambda) := \log E[e^{\lambda X_1}]$ is a convex function with

$$f(0) = 0, f'(0) = E[X_1], f''(0) = \frac{E[X_1^2 e^{\lambda X_1}] E[e^{\lambda X_1}] - E[X_1 e^{\lambda X_1}]^2}{E[e^{\lambda X_1}]^2} \geq 0.$$

Also, the rate function is convex and satisfies

$$I(0) = 0, I'(0) = 0, I''(0) = \frac{\sigma^2}{2} = \frac{\text{Var}(X_1)}{2}.$$

We remark that if $I(a) < \infty$, then \bar{S}_n "visits" $(a - \varepsilon, a + \varepsilon)$ with positive probability.

Theorem 6.6. The moment generating function of a random variable X is defined to be $M(\lambda) := E[e^{\lambda X}]$. Define

$$I(a) := \sup_{\lambda \in \mathbb{R}} \{a\lambda - \log M(\lambda)\}.$$

Then

$$(1) \mathbb{P}(\bar{S}_n \geq a) \leq e^{-nI(a)}, \text{ for } a > \mu;$$

$$(2) \mathbb{P}(\bar{S}_n \leq a) \leq e^{-nI(a)}, \text{ for } a < \mu;$$

(3) We have that

$$\mathbb{P}(|\bar{S}_n - a| < \varepsilon) \geq \left(\frac{1 - F(a)}{n\varepsilon^2} \right) e^{-n(I(a) + \varepsilon G(a))},$$

where the functions F, G are explicit in terms of μ, I, M .

Theorem 6.7 (WLLN, V3). *If $(X_i)_{i \geq 1}$ are uncorrelated with mean μ such that $\sigma^2 = \sup_{i \geq 1} \text{Var}(X_i) < \infty$, then $\bar{S}_n \xrightarrow{P, L^2} \mu$ as $n \rightarrow \infty$.*

Theorem 6.8 (WLLN, First Improvement, V4). *If $(X_i)_{i \geq 1}$ are iid, $\mu = E[X_i]$, and $E[X_1^4] < \infty$, then $\bar{S}_n \xrightarrow{P, L^4, a.s.} \mu$ as $n \rightarrow \infty$.*

Proof. This proof is instructive. We assume that we have $\mu = 0$, as otherwise we can take $\tilde{X}_i := X_i - \mu$ and notice that $\|\tilde{X}_i\|_{L^4} \leq \mu + \|X_i\|_{L^4} < \infty$. Next, we can expand

$$E[\bar{S}_n^4] = \frac{1}{n^4} E[(X_1 + \cdots + X_n)^4] = \frac{1}{n^4} \sum_{i,j,k,\ell=1}^n E[X_i X_j X_k X_\ell].$$

Observe that if one of the indices i, j, k, ℓ appears by itself in the previous expansion, then $E[X_i X_j X_k X_\ell] = 0$ (by independence). The surviving terms are of the forms $(ijk\hat{\ell}), (ij\hat{k}\ell), (\hat{i}jk\ell)$ and $i = j = k = \ell$. In the first three cases, we obtain terms of $E[X_1^2]^2$, and in the last case we obtain $E[X_1^4]$. This implies that

$$E[\bar{S}_n^4] = \frac{1}{n^4} (3n(n-1)E[X_1^2]^2 + nE[X_1^4]).$$

Now since $\text{Var}(X_1^2) = E[X_1^4] - E[X_1^2]^2 \geq 0$, we have that

$$E[\bar{S}_n^4] \leq \frac{1}{n^4} (3n(n-1) + n)E[X_1^4] \leq \frac{3E[X_1^4]}{n^2}.$$

Then this implies that

$$\mathbb{P}(|\bar{S}_n| \geq \varepsilon) = \mathbb{P}(\bar{S}_n^4 \geq \varepsilon^4) \leq \frac{3E[X_1^4]}{n^2 \varepsilon^4} \rightarrow 0,$$

as $n \rightarrow \infty$. And also, by Borel-Cantelli, we have that $\bar{S}_n \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$. \square

Remark 6.9 (Observations and Extensions). If (X_i) are independent, $E[X_i] = 0$ for all $i \geq 1$, and $\sup_{i \geq 1} E[X_i^4] < \infty$, then $\bar{S}_n \xrightarrow{P, L^4, a.s.} 0$. If $(X_i)_{i \geq 1}$ are independent, $E[X_i] = 0 \forall i$, and $\sup_{i \geq 1} E[X_i^{2k}] < \infty$, then $\bar{S}_n \xrightarrow{P, L^{2k}, a.s.} 0$. In fact, we can show that $\mathbb{P}(|\bar{S}_n| \geq \varepsilon) \leq \frac{C_k}{(n\varepsilon^2)^k}$, where the constant C_k depends only on the fixed $k \geq 1$.

6.3 The strong law of large numbers

Theorem 6.10 (SLLN). *Suppose that X_1, \dots, X_n are iid. Then $\bar{S}_n \xrightarrow{a.s.} \mu$ as $n \rightarrow \infty$ if and only if $E[|X_1|] < \infty$ and $E[|X_1|] = \mu$.*

Remark 6.11 (Finite Expectations and Forward Direction Proof Steps). Recall that

$$E[|X|] < \infty \iff \sum_{n \geq 1} \mathbb{P}[|X| > n] < \infty.$$

This is because we know that

$$E[|X|] = \int_0^\infty \mathbb{P}[|X| \geq x] dx.$$

And as we can see by taking unit-width rectangles, i.e., to show that

$$f(1) + f(2) + \cdots \leq \int_0^\infty f(x) dx \leq f(0) + f(1) + f(2) + \cdots,$$

we have that when f is non-increasing (for all large enough $x \gg 1$):

$$\int_0^\infty f(x)dx < \infty \iff \sum_{n \geq 1} f(n) < \infty.$$

Now, secondly, we also observe that if $\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{a.s.} \mu$, then $\frac{X_n}{n} = \frac{S_n - S_{n-1}}{n} \xrightarrow{a.s.} 0$. To see this we look at the limiting behavior of $\frac{X_n}{n} = \frac{S_n}{n} - \left(\frac{n-1}{n}\right) \frac{S_{n-1}}{n-1} \rightarrow 0$. So we conclude that $\frac{X_n}{n} \xrightarrow{a.s.} 0$.

Claim: This observation implies that X_1 is integrable.

Proof. Suppose that in fact X_1 is not integrable. Then we have that $\sum_{n \geq 1} \mathbb{P}[|X_1| > n] = +\infty$. So in fact since the X_i 's are independent, by Borel-Cantelli (II), $\mathbb{P}[\limsup_{n \rightarrow \infty} \{|X_n| > n\}] = 1$. But this contradicts our fact proven above that $\frac{X_n}{n} \rightarrow 0 \neq 1$. \square

Finally, in summary if $E[|X_1|] < \infty$, we have proved that $\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mu$, with $\mu = E[|X_1|]$. Also, though we have that $\frac{S_n}{n} \xrightarrow{a.s.} \mu_0$, so we obtain that in fact this $\mu_0 = E[|X_1|]$. This concludes the proof of the forward direction of the SLLN.

Lemma 6.12 (Real Analysis Fact). *If $\sum_{n \geq 1} \frac{x_n}{n} < \infty$, then immediately $\frac{x_1 + \dots + x_n}{n} \rightarrow 0$.*

Proof. By hypothesis, we have that $S_n \xrightarrow{S}$ as $n \rightarrow \infty$. Now if $S_n := \sum_{i=1}^n \frac{x_i}{i}$, then $x_n = (S_n - S_{n-1}) \cdot n$. So

$$\begin{aligned} \frac{x_1 + \dots + x_n}{n} &= \frac{1}{n} (S_1 - S_0 + 2(S_2 - S_1) + \dots + n(S_n - S_{n-1})) \\ &= \frac{1}{n} (S_0 - S_1 - S_2 - \dots - S_{n-1} + nS_n) \\ &= S_n - \frac{S_0 + \dots + S_{n-1}}{n} \rightarrow S := 0. \end{aligned} \quad \square$$

Theorem 6.13 (Kolmogorov). *If Y_n is a sequence of independent random variables such that $\sum_{n \geq 1} \text{Var}(Y_n) < \infty$, then $\sum_{n \geq 1} Y_n$ is a.s. convergent. If $Y_n := \frac{X_n}{n}$ satisfies $\sum_{n \geq 1} \frac{\text{Var}(X_n)}{n^2} < \infty$, then $\sum_{n \geq 1} \frac{X_n - E[X_n]}{n} < \infty$. As a consequence, if X_1, \dots, X_n are iid with mean μ and finite variance, then $S_n \xrightarrow{a.s.} \mu$.*

As a comment, if $\text{Var}(X_1) < \infty$, we have already proved that

$$\mathbb{P}[|S_n - \mu| > \varepsilon] \leq \frac{\text{Var}(X_1)}{n\varepsilon^2},$$

but in this case it's NOT Borel-Cantelli that provides the a.s. convergence.

Remark 6.14. To finish the sketch of the reverse direction of the SLLN, we next sketch the last remaining key ideas to this part of the proof. In particular, we can 1) use Borel-Cantelli to show the following:

$$\sum_{n \geq 1} \mathbb{P}[Y_n \neq X_n] = \sum_{n \geq 1} \mathbb{P}[|X_n| > n] = \sum_{n \geq 1} \mathbb{P}[|X_1| > n] \implies \mathbb{P}[\limsup_{n \rightarrow \infty} \{X_n \neq Y_n\}] = 0,$$

i.e., for all large enough n , $X_n = Y_n$ (a.s.) – so that

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow{a.s.} \mu \iff \frac{Y_1 + \dots + Y_n}{n} \xrightarrow{a.s.} \mu.$$

Then 2) we can use Kolmogorov's theorem on the variance to show that $\frac{Y_1 + \dots + Y_n}{n} \xrightarrow{a.s.} \mu$.

Corollary 6.15 (Consequences of the SLLN). *If X_1, \dots, X_n are iid, then for any $f : \mathbb{R} \rightarrow \mathbb{R}$ which is measurable and bounded, we have that $\sum_{i=1}^n \frac{f(X_i)}{n} \rightarrow E[f(X_1)]$. For example, suppose that $f(x) := \mathbf{X}_{[a,b]}(x)$. Then $\frac{1}{n} \sum_{i=1}^n \mathbf{X}_{[a,b]}(X_i) \rightarrow E[\mathbf{X}_{[a,b]}(X_1)] = \mathbb{P}[a \leq X_1 \leq b]$, where the former (non-limiting) sum corresponds to the relative frequency of X_i in $[a, b]$. Now for the special case where $a = -\infty$ and $b = x$, we obtain that*

$$\frac{1}{n} \sum_{i=1}^n \mathbf{X}_{(-\infty, x]}(X_i) \longrightarrow F_{X_1}(x) \equiv \mathbb{P}[X_1 \leq x].$$

Example 6.16 (Spring 2018, #2). Suppose that f is a continuous function on $[0, 1]$. Use the Law of Large Numbers to prove that

$$\lim_{n \rightarrow \infty} \int_0^1 \cdots \int_0^1 f((x_1 \cdots x_n)^{1/n}) dx_1 \cdots dx_n = f\left(\frac{1}{e}\right).$$

(*Hints: Use continuity and dominated convergence in the end.*)

Example 6.17 (Fall 2018, #1). Use the SLLN to find the following limit:

$$\lim_{n \rightarrow \infty} \int_0^1 \cdots \int_0^1 \frac{x_1^2 + \cdots + x_n^2}{x_1 + \cdots + x_n} dx_1 \cdots dx_n.$$

Example 6.18 (Spring 2017, #3). Let X_1, X_2, \dots be iid random variables uniformly distributed on $[0, 1]$. Show that with probability 1,

$$\lim_{n \rightarrow \infty} (X_1 \cdots X_n)^{1/n},$$

exists and compute its value.

Example 6.19 (Spring 2015, #8). Let $(X_n)_{n \geq 1}$ be an iid sequence of positive random variables such that $E[X_1] < \infty$. Let

$$N_t := \sup\{n : X_1 + \cdots + X_n \leq t\}.$$

Show that

$$\frac{N_t}{t} \xrightarrow[n \rightarrow \infty]{a.s.} \frac{1}{E[X_1]},$$

where the convergence is in the almost sure sense.

Example 6.20 (Fall 2019, #6). Assume $(X_n)_{n \geq 1}$ is a iid sequence of positive random variables. Show that

$$\frac{X_1 + X_2^2 + \cdots + X_n^n}{n} \xrightarrow[n \rightarrow \infty]{a.s.} 1 \iff X_1 = 1, a.s.$$

6.4 Concentration inequalities

Theorem 6.21. *If X_1, \dots, X_n are iid such that $a \leq X_i \leq b$ for all i , then*

$$\mathbb{P}(|\bar{S}_n - \mu| > \varepsilon) \leq 2e^{-2n\varepsilon^2/(b-a)^2}.$$

Also, notice that

$$\text{Var}[X_i] \leq \frac{(b-a)^2}{4}.$$

Proof Notes. We require the following lemma: If $a \leq X \leq b$, then $\text{Var}(X) \leq \frac{(b-a)^2}{4}$. □

Proposition 6.22. A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is 1-Lipschitz if

$$|f(x_1, \dots, x_n) - f(y_1, \dots, y_n)| \leq \sqrt{\sum_{i=1}^n (x_i - y_i)^2}.$$

Suppose that X_1, \dots, X_n are iid, $X_1 \sim N(0, 1)$, and that f is 1-Lipschitz on \mathbb{R}^n such that f is differentiable and $\sup_{x \in \mathbb{R}^n} |\nabla f(x)| \leq 1$. Then

$$\mathbb{P}(|f(x_1, \dots, x_n) - E[f(x_1, \dots, x_n)]| > \varepsilon) \leq e^{-\varepsilon^2/2}.$$

Example 6.23 (Cauchy-Schwarz and 1-Lipschitz Functions). (**Important.**) An example of a 1-Lipschitz function f is given by $f(x_1, \dots, x_n) := \frac{x_1 + \dots + x_n}{\sqrt{n}}$. To show this, we can pull a trick out of our hat in the form of "forcing" the Cauchy-Schwarz inequality in the following form:

$$\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n a_i\right)^2 \leq \frac{1}{n} \left(\sum_{i=1}^n 1^2\right) \left(\sum_{i=1}^n a_i^2\right) = \sum_{i=1}^n a_i^2.$$

In our case, we have that

$$|f(\vec{x}) - f(\vec{y})| \leq \sum_{i=1}^n |x_i - y_i| \implies \frac{1}{\sqrt{n}} \sum_{i=1}^n |x_i - y_i| \leq \sqrt{\sum_{i=1}^n |x_i - y_i|^2}.$$

Example 6.24. If X_1, \dots, X_n are iid with $X_1 \sim N(0, 1)$, we get that for $f(x_1, \dots, x_n) := \frac{1}{\sqrt{n}}(x_1 + \dots + x_n)$:

$$\begin{aligned} \mathbb{P}\left(\left|\frac{1}{\sqrt{n}} \sum_i X_i - E\left[\frac{1}{\sqrt{n}} \sum_i X_i\right]\right| > \lambda\right) &\leq 2e^{-\lambda^2/2} \\ \mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| > \sqrt{n}\lambda\right) &\leq 2e^{-\lambda^2/2}, \quad \forall \lambda > 0. \end{aligned}$$

And if we take $\lambda := \sqrt{n}\varepsilon$, then

$$\mathbb{P}\left(\frac{1}{n} \left|\sum_{i=1}^n X_i\right| > \varepsilon\right) \leq 2e^{-n\varepsilon^2/2}.$$

Example 6.25 (Spring 2018, #6). Let $(X_n)_{n \geq 1}$ be a sequence of iid random variables with

$$\mathbb{P}(X_n = 1) = \frac{1}{2} = \mathbb{P}(X_n = -1).$$

Let $(Y_n)_{n \geq 1}$ be a bounded sequence of random variables such that $\mathbb{P}(Y_n \neq X_n) \leq e^{-n}$. Show that

$$\frac{1}{n} E[(Y_1 + \dots + Y_n)^2] \longrightarrow 1,$$

as $n \rightarrow \infty$.

Example 6.26 (Spring 2017, #2; Spring 2016, #1). Let X be a random variable with mean zero and finite variance σ^2 . Prove that for every $c > 0$,

$$\mathbb{P}(X > c) \leq \frac{\sigma^2}{\sigma^2 + c^2}.$$

(HINT: Combine the inequality $E[c - X] \leq E[(c - X)\mathbf{1}_{X < c}]$ with the Cauchy-Schwarz inequality.)

(HINT: Write $c - X = (c - X)_+ - (c - X)_-$ and then use Cauchy-Schwarz.) (*)

7 Normal distributions

7.1 One-dimensional case

Definition 7.1 (Normal and Standard Normal Distributions). We say that $Z \sim N(0, 1)$ and $X \sim N(\mu, \sigma^2)$ if their probability density functions satisfy

$$f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathbb{R}$$

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}, \quad x \in \mathbb{R}.$$

One key important fact relating these two normal distributions is that if $X \sim N(\mu, \sigma^2)$, then the random variable $Z := \frac{X-\mu}{\sigma} \sim N(0, 1)$. In particular, given this definition, $F_Z(z) = F_X(\sigma z + \mu)$ and $f_Z(z) = \sigma \cdot f_X(\sigma z + \mu)$.

Proposition 7.2 (Properties of the Standard Normal). *We have the following properties of a standard normal random variable $Z \sim N(0, 1)$:*

- (1) $E[Z] = 0$ and $\text{Var}(Z) = 1$;
- (2) $M_Z(\lambda) = E[e^{\lambda Z}] = e^{\lambda^2/2}$ for all $\lambda \in \mathbb{C}$.

Proof of (1). First, we compute that

$$E[Z] = \int_{-\infty}^{\infty} z f_Z(z) dz = \int_{-\infty}^{\infty} \frac{z}{\sqrt{2\pi}} e^{-z^2/2} dz = 0,$$

since the right-hand-side integral is of an odd function over a symmetric interval. Next, we can compute by repeated integration by parts that

$$\begin{aligned} \text{Var}(Z) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-z^2/2} dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z \left(-e^{-z^2/2} \right) dz \\ &= z e^{-z^2/2} \Big|_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} dz \\ &= 1. \end{aligned}$$

□

Proof of (2). We have that

$$\begin{aligned} M_Z(\lambda) &= E[e^{\lambda Z}] = \int_{-\infty}^{\infty} \frac{e^{\lambda z}}{\sqrt{2\pi}} e^{-z^2/2} dz \\ &= \frac{e^{\lambda^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z^2 - 2\lambda z + \lambda^2)} dz \\ &= e^{\lambda^2/2}. \end{aligned}$$

□

Corollary 7.3. *If $X \sim N(\mu, \sigma^2)$, then its moment generating function is given by $M_X(\lambda) = e^{\lambda\mu + (\lambda\sigma)^2/2}$.*

Proof. We can write $X = \mu + \sigma Z$, where $Z \sim N(0, 1)$. This implies that

$$M_X(\lambda) = E[e^{\lambda X}] = E[e^{\lambda(\mu + \sigma Z)}] = e^{\lambda\mu} E[e^{\lambda\sigma Z}] = e^{\lambda\mu + \lambda^2\sigma^2/2}. \quad \square$$

Note that the moments of a normal $Z \sim N(\mu, \sigma^2)$ are given by

$$E[Z^n] = \begin{cases} 0, & \text{if } n \text{ is odd;} \\ \sigma^n \cdot (n-1)!! = \sigma^n(n-1)(n-3)\cdots 3 \cdot 1, & \text{if } n \text{ is even.} \end{cases}$$

Example 7.4 (Fall 2018, #3). Let $(Z_n)_{n \geq 1}$ be iid standard normal random variables, and let $(a_n)_{n \geq 1} \subset \mathbb{R}$. Prove that $\sum_{n \geq 1} a_n Z_n^2 < \infty \iff \sum_{n \geq 1} a_n < \infty$.

Example 7.5 (Spring 2016, #6). Let X_1, X_2, \dots be iid standard normal random variables, and for $x \in (-1, 1)$, set

$$Y := \sum_{n \geq 1} x^n X_n.$$

Show that the sum defining Y converges and find its distribution.

(See this example showing the sum is convergent using Borel-Cantelli.)

7.2 Multidimensional normal distributions

Definition 7.6. The random variable $Z := (Z_1, \dots, Z_d)$ is a *standard multidimensional normal vector* if Z_1, \dots, Z_d are iid with $Z_1 \sim N(0, 1)$. We write that $X = (X_1, \dots, X_d) \sim N(\mu, C)$ with $\mu \in \mathbb{R}^d$ and covariance matrix C if ¹

$$f_X(x_1, \dots, x_d) = \frac{1}{\sqrt{(2\pi)^d (\det C)^{1/2}}} e^{-\langle C^{-1}(x-\mu), x-\mu \rangle / 2}.$$

One particularly nice property relating these two types of multidimensional normal variables is that $X \sim N(\mu, C) \iff Z = C^{-1/2}(X - \mu) \sim N(0, \text{Id}) \iff X = \mu + C^{1/2}Z$ with $Z \sim N(0, \text{Id})$.

Exercise 7.7. Show that if $X = (X_1, \dots, X_n) \sim N(\mu, C)$, then $C_{ij} = \text{Cov}(X_i, X_j)$ and $E[X_i] = \mu_i$.

Proposition 7.8 (Properties and Transformations). *We have the following properties and transformation results relating multidimensional normal vectors:*

- (1) $X \sim N(\mu, C) \iff X = \mu + C^{1/2}Z$ where $Z = (Z_1, \dots, Z_d)$ for Z_i iid and $Z_1 \sim N(0, 1)$.
- (2) If $X \sim N(\mu, C)$, $A: \mathbb{R}^d \rightarrow \mathbb{R}^m$ is a linear map, then $Y := AX \sim N(A\mu, ACA^T)$.
- (3) If $X \sim N(\mu, C)$, then for any $a_1, \dots, a_d \in \mathbb{R}$, $\sum_{i=1}^d a_i X_i \sim N\left(\sum_{i=1}^d a_i \mu_i, \sum_{i,j=1}^d C_{ij} a_i a_j\right)$.
- (4) If $Z \sim N(0, \text{Id})$ and U is an orthogonal matrix, then $\tilde{Z} := UZ \sim N(0, \text{Id})$.

Sketch of (2). We first cite the following theorem: If $M_X(\lambda) = M_Y(\lambda)$ for all $\|\lambda\|_2 < \varepsilon$ ($\forall \varepsilon > 0$), then $X \sim Y$ have the same multidimensional distribution.

Step 1: Notice that

$$M_X(\lambda) = E[e^{\langle \lambda, X \rangle}] = E[e^{\sum_i \lambda_i X_i}].$$

¹Note that if $C \in \mathbb{C}^{d \times d}$ is symmetric, $C \geq 0$, then $C = UDU^T$ where U is orthogonal and D is a diagonal matrix of eigenvalues of C . In particular, this shows that $C^{1/2} = UD^{1/2}U^T$ in the computations from the last definitions.

Step 2: We compute that

$$\begin{aligned}
M_{AX}(\lambda) &= E \left[e^{\langle \lambda, AX \rangle_{\mathbb{R}^m}} \right] = E \left[e^{\langle A^T \lambda, X \rangle_{\mathbb{R}^d}} \right] \\
&= e^{\langle A^T \lambda, \mu \rangle + \langle CA^T \lambda, A^T \lambda \rangle} \\
&= e^{\langle \lambda, A\mu \rangle + \langle ACA^T \lambda, \lambda \rangle} \\
&= M_{N(A\mu, ACA^T)}(\lambda). \quad \square
\end{aligned}$$

Example 7.9. Let $X := [X_1, X_2]^T \sim N(\mu, C)$ where $X = \mu + C^{1/2}Z$ for $Z \sim N(0, \text{Id})$. Suppose that

$$C = \begin{bmatrix} \text{Cov}_{X_1} & 0 \\ 0 & \text{Cov}_{X_2} \end{bmatrix},$$

where $\text{Cov}_{X_1, X_2} = 0 \implies X_1, X_2$ are normal and independent. Then

$$X = \mu + C^{1/2} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \mu + \begin{bmatrix} \text{Cov}_{X_1}^{1/2} Z_1 \\ \text{Cov}_{X_2}^{1/2} Z_2 \end{bmatrix},$$

so that $X_i = \mu_i \text{Cov}_{X_i}^{1/2} Z_i$ are independent with the Z_i independent.

Theorem 7.10 (Concentration Inequality). *If Z_1, \dots, Z_d are iid with $Z_1 \sim N(0, 1)$, then for any 1-Lipschitz function $f : \mathbb{R}^d \rightarrow \mathbb{R}$, we have that*

$$\mathbb{P}(|f(Z_1, \dots, Z_d) - E[f(Z_1, \dots, Z_d)]|) \leq C_1 \cdot e^{-C_2 \varepsilon^2},$$

where the sharp second constant is given by $C_2 = 1/2$.

Proof Components. We cite *Jensen's inequality* which states that for ϕ convex: $E[\phi(X)] \geq \phi(E[X])$. We also use the identity that $E[e^{\lambda f(Z)}]E[e^{-\lambda f(Z)}] = E[e^{\lambda(f(Z_1) - f(Z_2))}]$. \square

Example 7.11 (Fall 2016, #4). Let (X, Y) be a normal vector in \mathbb{R}^2 with mean zero and covariance matrix:

$$\Sigma = \begin{bmatrix} 5 & 1 \\ 1 & 10 \end{bmatrix}.$$

Find $E[X^2 Y^2]$.

Example 7.12 (Spring 2016, #2). Let $X = (X_1, X_2)$ be a Gaussian vector with zero mean and covariance matrix

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},$$

where $|\rho| < 1$. Find a matrix A such that $X = AZ$ where Z is a standard normal vector, and derive the characteristic function of X as a function of ρ .

Example 7.13 (Spring 2015, #3). Assume that (X, Y) is a joint normal vector with $E[X] = E[Y] = 0$. Show that

$$E[X^2 Y^2] \geq E[X^2]E[Y^2],$$

with equality if and only if X, Y are independent.

Remark 7.14 (Notes on the Examples). If we assume that (X, Y) is a joint normal vector with mean $E[X] = E[Y] = 0$ and variance σ^2 , then the covariance matrix for the pair can be written as

$$C = \begin{bmatrix} \sigma^2 & \rho \\ \rho & \sigma^2 \end{bmatrix}.$$

This implies a formula for the characteristic function of (X, Y) given by

$$f_{X,Y}(\xi, \eta) = E[e^{i\xi X + i\eta Y}] = \exp\left(-\frac{(\xi\sigma)^2 + (\eta\sigma)^2 + 2\rho\xi\eta}{2}\right).$$

One usage of this characteristic function would be to compute the expectations

$$E[X^2Y^2] = \frac{1}{i^4} \frac{d^4}{d\xi^2 d\eta^2} f_{X,Y}(\xi, \eta) \Big|_{\xi=\eta=0}.$$

8 Central limit theorem

Theorem 8.1 (Central Limit Theorem, V1). Let X_1, \dots, X_n, \dots be independent iid random variables with mean μ and finite variance σ^2 . If $-\infty < a < b < \infty$, then

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(a \leq \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \leq b \right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx,$$

where $\sigma \equiv \sqrt{\text{Var}(X_1)}$. This means that there is convergence of the summation of rvs to a normal random variable with mean $\mu = 0$, AND, the convergence is weak to $N(0, 1)$.

Theorem 8.2 (Central Limit Theorem, V2). Let X_1, \dots, X_n, \dots be independent iid random variables with mean μ and finite variance σ^2 . Let $\bar{S}_n := \frac{X_1 + X_2 + \dots + X_n}{n}$. Then

$$\sqrt{n} (\bar{S}_n - \mu) \xrightarrow{\text{dist}} N(0, \sigma^2),$$

as $n \rightarrow \infty$.

Example 8.3 (Spring 2018, #4). Let X_1, \dots, X_n be iid random variables with mean μ and variance $\sigma^2 < \infty$. Let f be a function continuously differentiable at the point μ . Prove that the sequence of random variables

$$\sqrt{n} \left(f \left(\frac{X_1 + \dots + X_n}{n} \right) - f(\mu) \right),$$

converges in distribution to a normal random variable. What is the mean and the variance of the limit? (HINT: Use the first-degree Taylor approximation of f about μ .)

Proof. If we define $Y_n := \sqrt{n} \left(\frac{X_1 + \dots + X_n}{n} - \mu \right)$, then $Y_n \xrightarrow{\text{dist}} Y$, where $Y \sim N(0, \sigma^2)$ (by the Central Limit Theorem). Also, by the SLLN $Y_n/\sqrt{n} \xrightarrow{\mathbb{P}} 0$. \square

Example 8.4 (Spring 2018, #5). Let X_1, \dots, X_n, \dots be iid random variables with $E[X_1] = 0$ and $\text{Var}(X_1) = 1$. Define $S_n := X_1 + \dots + X_n$. Prove that

$$\limsup_{n \rightarrow \infty} \frac{S_n}{\sqrt{n}} = +\infty.$$

(Uses Kolmogorov's 0-1 law.)

Example 8.5 (Spring 2017, #6). Let (X_i) be iid random variables which are uniformly distributed on $[0, 2]$. Let $S_n := X_1 + \dots + X_n$. Show that

$$\frac{3\sqrt{3}}{2} n^{1/6} \left(\sqrt[3]{S_n} - \sqrt[3]{n} \right) \implies {}^w Z,$$

where Z is a standard normal random variable.

Example 8.6 (Spring 2016, #7). Let X_1, X_2, \dots be independent random variables such that $X_n \sim \text{Bernoulli}(p_n)$ for some $p_n > 0$. Show that if $np_n(1 - p_n) \rightarrow \infty$, then

$$\frac{X_n - np_n}{\sqrt{np_n(1 - p_n)}} \implies {}^w N(0, 1).$$

(Uses Lindeberg's CLT.)

Example 8.7 (Spring 2015, #2). Assume that X_1, X_2, \dots is a sequence of iid random variables such that for some $\alpha < 1/2$,

$$\frac{X_1 + \dots + X_n}{n^\alpha} \xrightarrow[n \rightarrow \infty]{a.s.} m,$$

for some real number m . Show that almost surely $X_i = 0$.

(Try to show that both X_i, X_i^2 are integrable, then derive a contradiction.)

Theorem 8.8 (Berry-Esseen). Suppose that $(X_n)_{n \geq 1}$ are iid random variables with $\mathbb{E}X_1 = 0$, $\mathbb{E}X_1^2 = \sigma^2 > 0$ and $\mathbb{E}|X_1|^3 = \rho < \infty$. Set $Y_n := \frac{X_1 + \dots + X_n}{n}$ and let F_n be the CDF of $\sqrt{n} \cdot Y_n / \sigma$. Then for all x, n , if $\Phi(x)$ is the distribution function of the standard normal, we have that there is an absolute constant $C > 0$ such that

$$|F_n(x) - \Phi(x)| \leq \frac{C\rho}{\sigma^3 \sqrt{n}}.$$

8.1 Extra problems with proofs and/or hints

Exercise 8.9. Let $(X_n)_{n \geq 1}$ be iid random variables with $X_n \geq 1$, $\mathbb{E}X_n = 1$ and $\text{Var}(X_n) = \sigma^2 \in (0, \infty)$. Show that

$$2(\sqrt{S_n} - \sqrt{n}) \implies N(0, 1),$$

as $n \rightarrow \infty$.

Proof Sketch. First, notice that

$$2(\sqrt{S_n} - \sqrt{n}) = \int_n^{S_n} \frac{dx}{\sqrt{x}}.$$

Now when $|S_n - n| \leq n^{2/3}$, an event that happens with probability tending to one as $n \rightarrow \infty$, we have that

$$\begin{aligned} \left| \int_n^{S_n} \frac{dx}{\sqrt{x}} - \frac{(S_n - n)}{\sqrt{n}} \right| &= \left| \int_n^{S_n} \left[\frac{1}{\sqrt{x}} - \frac{1}{\sqrt{n}} \right] dx \right| \\ &\leq n^{2/3} \left(\frac{1}{(n - n^{2/3})^{1/2}} - \frac{1}{n^{1/2}} \right) \\ &= n^{2/3} \int_{n-n^{2/3}}^n \frac{dx}{2x^{3/2}} \\ &\leq \frac{n^{4/3}}{2(n - n^{2/3})^{3/2}} \rightarrow 0, \end{aligned}$$

as $n \rightarrow \infty$. Now the only catch is to see that by the CLT, $\frac{S_n - n}{\sqrt{n}} \implies N(0, 1)$ as $n \rightarrow \infty$. \square

Exercise 8.10. Let $(X_n)_{n \geq 1}$ be iid random variables with $\mathbb{E}X_n = 0$ and $\mathbb{E}[X_n^2] = \sigma^2 \in (0, \infty)$. Show that

$$\frac{X_1 + \dots + X_n}{(X_1^2 + \dots + X_n^2)^{1/2}} \implies N(0, 1),$$

as $n \rightarrow \infty$.

Proof Sketch. By the WLLN, in probability we have that

$$\frac{X_1^2 + \dots + X_n^2}{n\sigma^2} \xrightarrow{\mathbb{P}} 1.$$

Since the function $y^{-1/2}$ is continuous at one, we also have that

$$\left(\frac{X_1^2 + \cdots + X_n^2}{n\sigma^2} \right)^{-1/2} \xrightarrow{\mathbb{P}} 1.$$

Now since convergence in probability implies convergence in distribution, and since by the CLT $\frac{X_1 + \cdots + X_n}{\sqrt{n\sigma}} \implies N(0, 1)$, we see the claim is true. \square

9 Useful inequalities and asymptotic estimates

9.1 Factorials, binomial coefficients and Stirling's inequality

We have that

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \left(1 + \frac{1}{12n} + \frac{1}{288n^2} + o(n^{-2})\right)$$

$$\binom{n}{k} \sim \sqrt{\frac{n}{2\pi k(n-k)}} \frac{n^n}{k^k (n-k)^{n-k}}.$$

We also have the inequalities that provide

$$\frac{n^k}{k^k} \leq \binom{n}{k} \leq \frac{n^k}{k!} < \left(\frac{ne}{k}\right)^k \quad (9.1a)$$

$$\sqrt{2\pi n} \cdot n^n e^{-n} \leq n! \leq e\sqrt{n} \cdot n^n e^{-n} \quad (9.1b)$$

$$\sqrt{2\pi n} \cdot n^n e^{-n} e^{\frac{1}{12n+1}} < n! < \sqrt{2\pi n} \cdot n^n e^{-n} e^{\frac{1}{12n}}. \quad (9.1c)$$

Inequalities on exponential functions and integrals

We have that as $n \rightarrow \infty$,

$$e^{-t^2/2} = \left(1 - \frac{t^2}{2n} + o(n^{-1})\right)^2.$$

For $w > 0$, we have

$$\int_w^\infty e^{-x^2/2} dx \leq \int_w^\infty \frac{x}{w} e^{-x^2/2} dx \leq \frac{e^{-w^2/2}}{w}$$

$$\int_w^\infty e^{-x^2/2} dx \geq \frac{we^{-w^2/2}}{1+w^2}.$$

Example 9.1. We have that for $n \geq 0$,

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

As a consequence,

$$\left| \mathbb{E} \left(e^{itX} - 1 + \frac{t^2}{2} X^2 \right) \right| \leq \mathbb{E} \left[\min \left(\frac{|t|^3 |X|^3}{6}, t^2 X^2 \right) \right]$$

$$= \mathbb{E} \left[\min \left(\frac{|t|^3 |X|^3}{6}, t^2 X^2 \right) (\mathbb{1}_{|X| < \varepsilon} + \mathbb{1}_{|X| \geq \varepsilon}) \right]$$

Other inequalities

- If $|z_i|, |w_j| < \vartheta \forall 1 \leq i, j \leq n$, then

$$\left| \prod_{m=1}^n z_m - \prod_{m=1}^n w_m \right| \leq \vartheta^{n-1} \sum_{m=1}^n |z_m - w_m|.$$

- $|e^{it} - 1|^2 \leq 4$.
- $\lim_{x \rightarrow 0} \frac{\log(1+x)}{x} = 1$.

10 Misc key facts, results, and other reminders

- If $(A_i)_{i \geq 1}$ are almost sure events, then $A = \bigcap_{i \geq 1} A_i$ is an almost sure event, i.e., if $\mathbb{P}(A_i) = 1$, then $\mathbb{P}(\bigcap_{i \geq 1} A_i) = 1$.
- If X is independent of itself, then X is almost surely constant. One consequence of this is that if X_i are independent and $a_1 X_1 + \dots + a_n X_n = 7$, then X_i is constant almost surely, as X_1 is equal to an expression independent of itself.
- $\{\sup_n X_n > t\} = \bigcup_{n \rightarrow 1} \{X_n > t\}$.
- For $\vartheta \in \mathbb{R}$, $|e^{\vartheta} - 1| \leq |\vartheta|$;
- **Variations of Inclusion-Exclusion:** For any two sets $E, F \subset \Omega$, we can write:

$$(A) \chi_{E \cup F} = \chi_E + \chi_F - \chi_{E \cap F}; \text{ and}$$

$$(B) \mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F),$$

where $\mathbb{P}(E \cap F) = \mathbb{P}(E) \cdot \mathbb{P}(F)$ iff E, F are independent. Also, for any three sets A, B, C : $\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C)$.

- Special sequences of sets constructions: $A_n := \{\omega : |X(\omega)| > n\}$;
- Special choices of functions: Take $f_n := n \cdot \chi_{[0, \frac{1}{n}]}$, or $f_n := \frac{1}{n} \cdot \chi_{[0, n]}$;
- $|a - b| = \frac{\max(a, b) - \min(a, b)}{2}$.
- $x + y = \min(x, y) + \max(x, y)$.
- $\chi_{[x, x+c]}(t) = \chi_{[t-c, t]}(x)$, which can be combined with an application of Fubini or Tonelli to swap the orders of integration in a double or multiple integral.
- Whenever see convergence in probability, try to extract a subsequence which converges almost surely.
- Whenever see (i) positive g_n ; and (ii) $g_n \rightarrow g$ a.e., immediately try Fatou's lemma.
- If cannot prove convergence in probability directly, think proof by contradiction.
- Let Δ denote the *symmetric difference of sets*: $A \Delta B = (A \setminus B) \cup (B \setminus A)$.
- $E \setminus F = (E \setminus A) \dot{\cup} (A \setminus F)$.
- $|A \setminus E|_e < |A \Delta E|_e$.
- $f(x) = \sum_{j=1}^n \chi_{E_j}(x) \implies \int_0^1 f = \sum_{j=1}^n |E_j|$.
- $CF \setminus CE = E \setminus F$ and $E_k = F_k \cup (E_k \setminus F_k)$.
- Let $E_k^{(j)} := \{x \in E_k : j - 1 \leq |x| < j\} = (E_k \cap \{x : |x| < j\}) \setminus \{x : |x| < j - 1\}$. Then $E_k^{(j)}$ is *bounded* and measurable. May need to assume that $|E_k| < \infty$ for some k (the other case is easier).
- $E \setminus \bigcap_{m \geq 1} F_m = \bigcup_{m \geq 1} E \setminus F_m$
- If $|E| < \infty$, then $\chi_E(x) \in L^1(\mathbb{R})$.

- Continuous on $E \implies$ measurable on E .
- $\mu(\{x : |f_n - f| + |g_n - g| \geq \varepsilon\}) \leq \mu(\{x : |f_n - f| \geq \varepsilon/2\}) + \mu(\{x : |g_n - g| \geq \varepsilon/2\})$
- $\inf_n x_n \leq \liminf_{n \rightarrow \infty} x_n \leq \limsup_{n \rightarrow \infty} x_n \leq \sup_n x_n$
where $\liminf_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} \inf_{m \geq n} x_m = \sup_{n \geq 0} \inf_{m \geq n} x_m$ and
 $\limsup_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} \sup_{m \geq n} x_m = \inf_{n \geq 0} \sup_{m \geq n} x_m$.
- $f(x) = \frac{1}{(x \log x)^{1/q}} \in L^q \notin L^p$ for $p < q$.
- Let $A_n := \{|f| > n\}$ and $A := \{|f| = \infty\}$. Then $A_{n+1} \subseteq A_n$ and $A = \bigcap_{n \geq 1} A_n$ so that $A_n \searrow A$. By continuity, $\lim_{n \rightarrow \infty} |A_n| = |A| = 0$ when f is finite a.e.
- f' integrable \implies for all $\varepsilon > 0$, $\exists \delta > 0$ such that for any measurable $A \subseteq \mathbb{R}$: $|A| < \delta \implies \int_A |f'| < \varepsilon$.
- $\liminf(s_n) + \liminf(t_n) \leq \liminf(s_n + t_n) \leq \limsup(s_n + t_n) \leq \limsup(s_n) + \limsup(t_n)$, for any sequences $\{s_n\}$ and $\{t_n\}$.
- $ab \leq \frac{1}{2} \left(\frac{a}{2} + b\right)^2$ and $ab \leq \frac{1}{4}(a + b)^2$.
- $1 + 1/2 + \dots + 1/n - (1 + 1/n) < \log(n) < 1 + 1/2 + \dots + 1/n$, so that $\frac{1+1/2+\dots+1/n}{\log(n)} \rightarrow 1$ as $n \rightarrow \infty$.
- The characteristic function of a random variable with mean 0 and variance 1 satisfies $f(\xi) = 1 - \xi^2/2 + o(\xi^2) = e^{-\xi^2/2 + o(\xi^2)}$.
- If X_1, X_2 are independent, then the characteristic function of $X_1 + X_2$ is the product of their characteristic functions: $\phi(u) = \phi_1(u)\phi_2(u)$.
- A consequence of Jensen's inequality for expectation: $E[X]E[1/X] \geq E[X]/E[X] = 1$.
- Convergence in probability implies convergence in distribution.
- If the characteristic function of X_n converges pointwise to that of X , then $X_n \implies X$.
- By the CLT, $Y_n := n^{1/2} \left(\frac{X_1 + \dots + X_n}{n} - \mu\right)$ converges in distribution to a normal random variables with mean 0 and variance σ^2 .
- By Cauchy-Schwarz,

$$\begin{aligned} E \left(\frac{S_n}{\sqrt{n}} - \frac{T_n}{\sqrt{n}} \right)^2 &= \frac{1}{n} E(S_n - T_n)^2 \chi_{A_n} + \frac{1}{n} E(S_n - T_n)^2 \chi_{A_n^c} \\ &\leq \frac{1}{n} \sqrt{E(S_n - T_n)^4 \mathbb{P}(A_n)} + \frac{1}{n} E(S_n - T_n)^2 \chi_{A_n^c}. \end{aligned}$$

- Using monotone convergence,

$$\sum_{n \geq 0} \mathbb{P}(X^2/c^2 \geq n) = E \left[\sum_{n \geq 0} \chi_{X^2/c^2 \geq n} \right] = E \left(\left\lfloor \frac{X^2}{c^2} \right\rfloor + 1 \right) \leq 1 + \frac{1}{c^2} E[X^2].$$

- In the Borel-Cantelli lemma, interpret $\mathbb{P}(\limsup A_n)$ as $\mathbb{P}(X \in A_n \text{ infinitely often})$.
- If $X_n \implies X$ (weak convergence), then $\liminf_{n \rightarrow \infty} \mathbb{P}(X_n > t) = \mathbb{P}(X > t)$.
- If $\{X_n\}_{n \geq 1}$ are uniformly integrable, then the sequence $\{E[X_n]\}_{n \geq 1}$ is bounded.
- We have: $\log(1 - t) = -t + O(t^2)$ for small t .
- $\mathbb{P}(P < X) \leq \sum_{r \text{ rational}} \mathbb{P}(Y \leq r < X)$.