MEASURE THEORY AND CENTRAL LIMIT THEOREM

DAVID WANG

ABSTRACT. This paper will show how measure theory is used in the construction of the classical central limit theorem, which states that the sample mean of independent, identically distributed random variables will converge in distribution to the standard normal distribution. While no knowledge of probability is necessary, certain results from measure theory will be assumed.

Contents

1. Introduction

In the field of probability, measure theory serves as a language to formally convey ideas that would otherwise exist only in the realm of intuition. In this paper, I aim to prove one such idea, the classical central limit theorem, using results developed in measure theory. Section 2 will serve as a primer on measure theory. Sections 3, 4, and 5 will develop results in probability spaces using the established knowledge from section 2. This will lead to section 6, where the proof of the classical central limit theorem will be given.

2. Measure Theory

In this section, we will recall terminology and some important results from measure theory. In the next section, we will see how these results are used as a baseline for probability theory.

Definition 2.1. A *σ-algebra* is a collection of subsets that is closed under complements, unions, and intersections. This means that if a subset E is in the σ-algebra Σ, then E^c must also be in Σ. Additionally, if $E_1, E_2 \in \Sigma$, then $E_1 \cap E_2 \in \Sigma$ and $E_1 \cup E_2 \in \Sigma$. Note that a σ-algebra generated by the subsets of a set must contain the set itself and the empty-set.

Definition 2.2. The *Borel algebra* $\mathcal{B}$ of $\mathbb{R}$ is the σ-algebra that is generated by the open sets of $\mathbb{R}$. Since a σ-algebra must be closed under complements, unions, and intersections, it follows that $\mathcal{B}$ contains all the closed sets, $\mathbb{R}$, $\emptyset$, and all the intersections.

The Borel algebra has an important application in probability when we discuss random variables and distribution functions.

Definition 2.3. A *measure* μ is an extended real valued function on the measure space (X, Σ) where $\mu : \Sigma \to [0, \infty]$, satisfying the following properties:

(1) $\mu(\emptyset) = 0$
(2) $\mu\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mu(A_j)$, if $A_j \in \Sigma$ where$\{A_j\}$ is a countable collection of pairwise, disjoint sets.

Definition 2.4. A measure μ on the measure space (X, Σ) is called σ*-finite* if the set X is the countable union of measurable sets with a finite measure. A measure μ is a *finite measure* if $\mu(X) < \infty$.

Theorem 2.5 (Properties of a measure)**.** *Consider a measure μ on the measure space (X, Σ). Then the following properties must hold true.*

1) Monotonicity: *If $E_1 \subset E_2$, then $\mu(E_1) \leq \mu(E_2)$.*

2) Subadditivity: *if $E \subset \bigcup_{j=1}^{\infty} E_j$, then $\mu(E) \leq \sum_{j=1}^{\infty} \mu(E_j)$.*

3) Continuity from above: *If $E_n \uparrow E$ where $E_i \subset E_{i+1}$, then*
$$\lim_{n\to\infty} \mu(E_n) = \mu(E).$$

4) Continuity from below: *If $E_n \downarrow E$ where $E_{i+1} \subset E_i$, then*
$$\lim_{n\to\infty} \mu(E_n) = \mu(E).$$

Proof. :

1) $E_2 = E_1 \cup (E_2 \cap E_1^c)$ where E_1 and $(E_2 \cap E_1^c)$ are pairwise, disjoint. Therefore, $\mu(E_2) = \mu(E_1) + \mu(E_2 \cap E_1^c) \geq \mu(E_1)$.

2) Let $E'_n = E_n \cap E$. Let $O_1 = E'_1$ and for $n > 1$, $O_n = E'_n \setminus \cup_{j=1}^{n-1} E_j$. Since $\cup_{j=1}^{n} O_j = E$ is a disjoint union, we can use monotonicity and the definition of a measure to show that
$$\mu(E) = \sum_{j=1}^{\infty} \mu(O_j) \leq \sum_{j=1}^{\infty} \mu(E_j).$$

3) Let $O_n = E_n \setminus E_{n-1}$. $\bigcup_{j=1}^{\infty} O_j = E$ is a disjoint union where
$$\bigcup_{j=1}^{n} O_j = E_n.$$
By definition,
$$\mu(E) = \sum_{j=1}^{\infty} \mu(O_j) = \lim_{n\to\infty} \sum_{j=1}^{n} \mu(O_j) = \lim_{n\to\infty} \mu(E_n).$$

4) From 3), $(E_1 \setminus E_n) \uparrow (E_1 \setminus E)$ implies that
$$\lim_{n\to\infty} \mu(E_1 \setminus E_n) = \mu(E_1 \setminus E).$$
Since $E \subset E_1$, we have $\mu(E_1 \setminus E) = \mu(E_1) - \mu(E)$. It follows that $\lim_{n\to\infty} \mu(E_n) = \mu(E)$.

□

Definition 2.6. Consider two measure spaces (X, Σ) and (Y, T). Then a function $f : (X, \Sigma) \to (Y, T)$ is a *measurable function* if $f^{-1}(B) \in \Sigma$ for every $B \in T$.

Definition 2.7. An *indicator function* on a measurable set E is defined as follows:
$$1_E(x) = \left\{ \begin{array}{ll} 1, & \text{for } x \in E, \\ 0, & \text{for } x \notin E. \end{array} \right\}$$

Definition 2.8. Let (X, Σ) be a measure space. Let $E_1, \ldots, E_n \in \Sigma$ be a sequence of pairwise disjoint sets and let $a_1, \ldots, a_n$ be a sequence of real or complex numbers. Then a *simple* function is a complex valued function $f : X \to \mathbb{C}$ of the form
$$f(x) = \sum_{k=1}^{n} a_k 1_{E_k}(x).$$

A consequence of simple functions in measure theory is that any non-negative, measurable function is the pointwise limit of some sequence of non-negative, measurable simple functions.

We now recall some important results from Lebesgue integration which will be used extensively in the upcoming sections.

Theorem 2.9 (Monotone Convergence Theorem)**.** *If $\{f_n : n \in \mathbb{N}\}$ is a monotone increasing sequence*
$$\ldots \le f_n \le f_{n+1} \le \ldots$$
of positive, extended, real-valued, measurable functions $f_n : X \to [0, \infty]$ and
$$f = \lim_{n\to\infty} f_n,$$
then
$$\lim_{n\to\infty} \int f_n(d\mu) = \int f(d\mu).$$

Lemma 2.10 (Fatou's Lemma)**.** *Let (X, Σ, μ) be a measure space and let $\{f_n\}$ be a sequence of non-negative measurable functions where $f_n : X \to [0, \infty]$. Then the function $\liminf_{n\to\infty} f_n$ is measurable and*
$$\int_X \liminf_{n\to\infty} f_n(d\mu) = \liminf_{n\to\infty} \int_X f_n(d\mu).$$

Proof. For each $n \in \mathbb{N}$, let
$$g_n := \inf_{k \ge n} f_n.$$

We know that $\{g_n\}$ is a monotone increasing sequence of functions that converges pointwise to the limit
$$\lim_{n\to\infty} g_n = \liminf_{n\to\infty} f_n,$$
which allows for monotone convergence theorem to apply:
$$\lim_{n\to\infty} \int g(d\mu) = \int \liminf_{n\to\infty} f_n(d\mu).$$

Moreover, $g_n \leq f_k$ for $k \geq n$ gives us

$$\int g_n(d\mu) \leq \inf_{k\geq n} \int f_k(d\mu).$$

By taking the limit $n \to \infty$ on both sides of the inequality and using the previous equality, we get the desired result. □

Theorem 2.11 (Dominated Convergence Theorem). *If $\{f_n\}$ is a sequence of measurable functions $f_n : X \to \mathbb{R}$ such that $f_n \to f$ point-wise, and $|f_n| \leq g$ where $g : X \to [0, \infty]$ is an integrable function, meaning $\int g(d\mu) < \infty$, then*

$$\lim_{n\to\infty} \int f_n(d\mu) = \int f(d\mu).$$

Proof. Since $g + f_n \geq 0$ for every $n \in \mathbb{N}$, Fatou's lemma implies that

$$\int (g + f)(d\mu) \leq \liminf_{n\to\infty} \int (g + f_n)(d\mu) \leq \int g(d\mu) + \liminf_{n\to\infty} \int f_n(d\mu).$$

By dropping $\int g(d\mu)$ from both sides, we get

$$\int f(d\mu) \leq \liminf_{n\to\infty} \int f_n(d\mu).$$

Similarly, we can apply Fatou's lemma again since $g - f_n \geq 0$. We have

$$\int (g - f)(d\mu) \leq \limsup_{n\to\infty} \int (g - f_n)(d\mu) \leq \int g(d\mu) - \limsup_{n\to\infty} \int f_n(d\mu),$$

which gives us

$$\int f(d\mu) \geq \limsup_{n\to\infty} \int f_n(d\mu).$$

The desired result follows. □

It should be noted that the dominated convergence theorem holds for sequences of complex valued functions if both the real and imaginary parts are measurable. Consider a complex valued function $f : X \to \mathbb{C}$ where $f = g + ih$. The function f is measurable if and only if both $g, h : X \to \mathbb{R}$ are measurable. f is integrable if and only if g, h are integrable. The integral is defined as

$$\int f(d\mu) = \int g(d\mu) + i \int h(d\mu).$$

The dominated convergence theorem has an important corollary that considers sequences of uniformly bounded functions, i.e. $|f_n| \leq M$.

Corollary 2.12 (Bounded Convergence Theorem). *Let $|f_n| \leq M$ be a uniformly bounded sequence of complex valued, measurable functions where $f_n \to f$ point-wise on a measure space $(X, \mathcal{A}, \mu)$ and $\mu(X) < \infty$. Then f is an integrable function and*

$$\lim_{n\to\infty} \int f_n(d\mu) = \int f(d\mu).$$

Proof. $|f_n|$ is uniformly bounded by a real number M for $x \in X$. Define $M = g(x)$ for all $x \in X$. g is an integrable function since it is constant on a finitely measurable set X. Therefore, the dominated convergence theorem applies. □

We conclude this section with a statement of Fubini's theorem. This is an especially important result when we consider its applications with characteristic functions.

Theorem 2.13 (Fubini's Theorem). *Suppose that $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ are σ-finite measure spaces. A measurable function $f : X \times Y \to \mathbb{C}$ is integrable if and only if either of the iterated integrals*

$$\int_X \left(\int_Y f(d\nu) \right) (d\mu), \qquad \int_Y \left(\int_X f(d\mu) \right) (d\nu)$$

is finite. If f is integrable, then

$$\int_{X \times Y} f(d\mu) \otimes (d\nu) = \int_X \left(\int_Y f(d\nu) \right) (d\mu) = \int_Y \left(\int_X f(d\mu) \right) (d\nu).$$

The proofs for Fubini's theorem and the monotone convergence theorem have been omitted and can be found in both [2] and [4].

3. Random Variables and Distribution Functions

In this section, we will define key concepts of probability, such as probability spaces, random variables, and independence. We will also see a few important results that will be used when discussing characteristic functions in section 4.

Definition 3.1. A *probability space* is a measure space with a total measure one and is often written as $(\Omega, \mathcal{F}, P)$.

- Ω is a set of *outcomes* with elements denoted by ω. Ω is often called the *sample space* in elementary probability.
- $\mathcal{F}$ is a σ-algebra of subsets of Ω where each subset is called an *event*.
- P, commonly known as the *probability*, is a function from $\mathcal{F}$ to [0,1] and such that if $E_1, E_2, \ldots \in \mathcal{F}$ are disjoint, then

$$P\left[\bigcup_{j=1}^{\infty} E_j \right] = \sum_{j=1}^{\infty} P[E_j].$$

Observe that P is actually a measure on Ω where $P(\Omega) = 1$. As a result, the probability measure P inherits the properties outlined in theorem 2.5.

Definition 3.2. A *random variable* X is a measurable function from the probability space $(\Omega, \mathcal{F}, P)$ to the reals, or any topological space for the more general case. Formally,

$$X : \Omega \to (-\infty, \infty)$$

such that for every Borel set B, we have

$$X^{-1}(B) = \{X \in B\} \in \mathcal{F}.$$

Here,

$$\{X \in B\} = \{\omega \in \Omega : X(\omega) \in B\}.$$

For a random variable X, we can identify a function μ_X for each element B in the Borel algebra $\mathcal{B}$ that maps to the interval [0,1]. We define the function as

$$\mu_X(B) = P\{X \in B\} = P[X^{-1}(B)].$$

This function is indeed a measure for the space $(\mathbb{R}, \mathcal{B})$ and is called the *distribution* of the random variable X.

Definition 3.3. A random variable X is considered to be a *discrete* random variable when the image of X is a countable set.

Definition 3.4. A random variable Y is considered to be *continuous* when the μ_Y gives the measure 0 for every singleton set. This includes countable sets.

Definition 3.5. The *distribution function* of a random variable X with distribution μ_X is defined as

$$F(x) := P(X \leq x) = \mu_X(-\infty, x].$$

Theorem 3.6. *The distribution function F satisfies the following properties:*

1) $\lim_{x \to -\infty} F(x) = 0$

2) $\lim_{x \to \infty} F(x) = 1$

3) F is non-decreasing

4) F is right continuous: $\lim_{\epsilon \to 0} F(x + \epsilon) = F(x)$ *for every x.*

Proof. :

1) Observe that as $x \to -\infty$, then $\{X \leq x\} \to \emptyset$. Applying the properties of a measure, we get the desired result.

2) Similarly, we can see that as $x \to \infty$, then $\{X \leq x\} \to \Omega$.

3) For any $x < y$, we get $\{X \leq x\} \subset \{X \leq y\}$. By properties of a measure, we can see that $F(x) = P(X \leq x) \leq P(X \leq y) = F(y)$ for any $x < y$.

4) Let $y = \epsilon + x$ for $\epsilon > 0$. Observe that as $\epsilon \to 0$, $\{X \leq y\} \downarrow \{X \leq x\}$, which allows continuity from below to apply.

□

We could also say that if a function F satisfies the above properties, then F is a distribution function for some random variable.

For some continuous random variables, there is a function $f = f_X : \mathbb{R} \to [0, \infty)$ where $P(a \leq X \leq b) = \int_a^b f(t)dt$. This function f is called the *density* of a random variable and if the function exists, then

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

If f is continuous on t, we can then use the fundamental theorem of calculus to show that

$$f(x) = F'(x).$$

The density f must also satisfy

$$\int_{-\infty}^{\infty} f(t)dt = 1.$$

Remark 3.7. If X is a random variable that maps from $(\Omega, \mathcal{F})$ to $\mathbb{R}$, then any function of the random variable $g(X)$ also maps from $(\Omega, \mathcal{F})$ to $\mathbb{R}$. Therefore, $g(X)$ is also a random variable.

Definition 3.8. If $X_1, X_2, \ldots, X_n$ are random variables, then we consider $(X_1, \ldots, X_n)$ to be a *random vector*, and thus a random variable taking values in $\mathbb{R}^n$. The *joint* *distribution* of the random vector $\mu_{X_1\dots,X_n}$ is a measure on the Borel sets B of $\mathbb{R}^n$ where

$$\mu_{X_1\dots,X_n} = P\{(X_1\dots,X_n) \in B\}.$$

Definition 3.9. We say a sequence of random variables X_n converges *in distribution* to a random variable X if

$$\lim_{n\to\infty} F_n(x) = F(x).$$

Definition 3.10. A random variable X has a normal distribution $\mathcal{N}\{\mu, \sigma^2\}$ if X has a density of

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x-\mu^2}{2\sigma^2}}.$$

X is said to have a standard normal distribution if $\mu = 0$ and $\sigma^2 = 1$. Its distribution function is denoted by

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

Note that μ and σ^2 represent the mean and the variance, both of which will be defined in the next section.

Remark 3.11. It is important to understand that the behavior of random variables defined using probability theory can be observed through the distribution functions. As such, when we say $X = Y$, we mean that X and Y can be represented using the same distribution functions, $F_X = F_Y$ or $P(X \leq x) = P(Y \leq x)$. This type of relation is called *equality in distribution*.

4. Expectation and Independence

In this section, we will define expectation, variance, and independence between random variables.

Definition 4.1. If X is a non-negative random variable, the *expectation* of X, denoted by $E[X]$, is

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

$E[X]$ is often called the *mean* and is usually denoted by μ, as we saw previously. The integral is defined using lebesgue integration. As such, all results from lebesgue integration apply for the expectation.

In the discrete case,

$$E[X] = \sum_{i=1}^{\infty} a_i P(X = a_i).$$

We can see that the expectation of the random variable only depends on the distribution rather than the probability space where it is defined. If the random variable X has a density of $f(x)$, then

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

The expectation has the following properties:

Theorem 4.2. *Assume random variables $X, Y \geq 0$ are given.*

(1) $E[X + Y] = E[X] + E[Y]$

(2) $E[aX + b] = aE[X] + b$
(3) *if* $X \leq Y$, *then* $E[X] \leq E[Y]$.

Lemma 4.3. *If a random variable* X *has a distribution of* μ_X, *and* g *is a Borel measurable function, then*

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

Proof. Without loss of generality, assume g is a non-negative measurable function. Then there exists an increasing sequence $\{g_n\}$ of non-negative, simple functions that converges pointwise to g. Following remark 3.7, this tells us that $g_n(X)$ is an increasing sequence of non-negative, simple random variables that converges pointwise to $g(X)$. For the simple functions g_n, we have

$$E[g_n(X)] = \int_{\mathbb{R}} g_n(x)(d\mu_X),$$

which follows directly from definition. By monotone convergence theorem,

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

For the general case, we write $g(X) = g^+(X) - g^-(X)$, where $g^+(X) = \max\{g(X), 0\}$ and $g^-(X) = \max\{-g(X), 0\}$. Because g^- and g^+ are non-negative measurable functions, $E[g^+(X)]$ and $E[g^-(X)]$ are finite. Following the previous case for non-negative functions,

$$\begin{aligned} E[g(X)] &= E[g^+(X)] - E[g^-(X)] \\ &= \int_{\mathbb{R}} g^+(x)(d\mu_X) - \int_{\mathbb{R}} g^-(x)(d\mu_X) \\ &= \int_{\mathbb{R}} g(x)(d\mu_X). \end{aligned}$$

□

This result allows us to prove important results about characteristic functions.

Definition 4.4. The *variance* of a random variable $Var(X)$ is defined as

$$Var(X) = E[(X - E[X])^2],$$

provided that the expectation exists. Note that

$$\begin{aligned} Var(X) &= E[(X - E[X])^2] \\ &= E[X^2 - 2XE[X] + E[X]^2] \\ &= E[X^2] - 2E[X]E[X] + E[X]^2 \\ &= E[X^2] - E[X]^2. \end{aligned}$$

It follows from definition that the following properties hold true:

(1) $Var(a) = 0$
(2) $Var(aX) = a^2 Var(X)$
(3) $Var(X + a) = Var(X)$.

The variance is often denoted by σ^2 where σ is the *standard deviation*.

Definition 4.5. Two random variables X, Y on the probability space $(\Omega, \mathcal{F}, P)$ are considered to be *independent* if for any $A, B \in \Omega$,

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

An equivalent definition of independence between random variables X, Y with distribution functions $\mu_X(x), \mu_Y(y)$ requires

$$\mu_{X,Y} = \mu_X(x)\mu_Y(y)$$

where $\mu_{X,Y}$ is the joint distribution.

Definition 4.6. Random variables $X_1, X_2, \ldots$ are *independent,identically distributed (i.i.d.)* random variables if they have identical distribution functions and are independent.

Proposition 4.7. *If $X_1, \ldots, X_n$ are independent random variables and g is a Borel measurable function, then $g(X_1), \ldots, g(X_n)$ are independent random variables.*

Theorem 4.8. *Suppose X, Y are independent random variables and have respective distributions of $\mu(x), \nu(y)$. If $h : \mathbb{R}^2 \to \mathbb{R}$ is a measurable function where $h \geq 0$, then*

$$E[h(X,Y)] = \int \int h(x,y)(d\mu)(d\nu).$$

Additionally, if $h(x,y) = f(x)g(y)$ where f, g are Borel measurable functions, then

$$E[f(X)g(Y)] = E[f(X)]E[g(Y)].$$

Proof. Following lemma 4.3 and Fubini's theorem, we get

$$E[h(X,Y)] = \int_{\mathbb{R}^2} h(x,y)(d\mu) \otimes (d\nu) = \int \int h(x,y)(d\mu)(d\nu).$$

For the second result, by applying lemma 4.3 twice,

$$\begin{aligned}
[f(X)g(Y)] &= \int \int f(x)g(y)(d\mu)(d\nu) \\
&= \int g(y) \int f(x)(d\mu)(d\nu) \\
&= \int g(y)E[f(X)](d\nu) \\
&= E[f(X)]E[g(Y)]
\end{aligned}$$

□

An immediate consequence of the second result is that if X, Y are random variables, then $E[XY] = E[X]E[Y]$ by defining $g(Y)f(X) = XY = h(X,Y)$. Random variables X, Y that satisfy $E[XY] = E[X]E[Y]$ are considered to be *orthogonal*. This implies that independent random variables must be orthogonal. However, not all orthogonal random variables are independent.

5. Characteristic Functions and The Inversion Formula

In this section, we will first define a characteristic function for any random variable. Once we define the inversion formula, we will see why finding the characteristic function is crucial for the proof of the central limit theorem.

Definition 5.1. A *characteristic function* of a random variable X on a probability space $(\Omega, \mathcal{F}, P)$ is a complex valued function ϕ, where

$$\phi(t) := E[e^{itX}] = E[\cos(tX)] + iE[\sin(tX)].$$

The second half of the equality suggests that there is no need to devise a new method of taking the expected value of a complex valued random variable. If Y is a complex valued random variable, then $E[Y] = E[Re(Y)] + iE[Im(Y)]$ where $Re(Y)$ is the real part and $Im(Y)$ is the imaginary part. The *modulus* of $y = a + bi$ is $|a + bi| = \sqrt{a^2 + b^2}$. The *complex conjugate* of $y = a + bi$ is $\overline{y} = a - bi$.

Definition 5.2. A *moment-generating function* $M_X(t)$ of a random variable X is

$$M_X(t) = e^{tX}$$

and can be used to find *moments* of a random variable, $E[X^k]$.

While $M_X(t)$ does not always exist for all values of t, we can use the relation $M_X(it) = \phi_X(t)$ when $M_X(t)$ does exist.

Theorem 5.3. *We note that all characteristic functions have the following properties:*

1) $\phi(0) = 1$

2) $\phi(-t) = \overline{\phi(t)}$

3) $|\phi(t)| = |E[e^{itX}]| \leq E[|e^{itX}|] = 1$

4) $|\phi(t + h) - \phi(t)| \leq E[|e^{ihX} - 1|]$, *which implies that* $\phi(t)$ *is uniformly continuous on* $(-\infty, \infty)$.

5) $E[e^{it(aX+b)}] = e^{itb}\phi(at)$

Proof. :

1) follows directly from the definition of a characteristic function.

2) Note that

$$\phi(-t) = E[\cos(-tX) + i\sin(-tX)] = E[\cos(tX)] - iE[\sin(tX)].$$

3) We have

$$\Big|\int_{-\infty}^{\infty} \cos(tx)+i\sin(tx)(d\mu_X)\Big| \leq \int_{-\infty}^{\infty} \Big|\cos(tx)+i\sin(tx)\Big|(d\mu_X) = E[|\cos(tX)+i\sin(tX)|],$$

and $|\cos(tX) + i\sin(tX)| = 1$.

4) $|\phi(t+h)-\phi(t)| = |E[e^{i(t+h)X} - e^{itX}]| \leq E[|e^{i(t+h)X} - e^{itX}|] = E[|e^{itX}(e^{ihX} - 1)|] = E[|e^{ihX} - 1|]$

5) $E[e^{it(aX+b)}] = E[e^{itb}e^{itaX}] = e^{itb}\phi(at)$.

∎

Theorem 5.4. *If X and Y are independent random variables and have characteristic functions of $\phi_X(t)$ and $\phi_Y(t)$, then $X+Y$ has a characteristic function of $\phi_{X,Y}(t) = \phi_X(t)\phi_Y(t)$.*

Proof. Since X, Y are independent, e^{itX} and e^{itY} are both independent random variables. By proposition 4.7 and theorem 4.8, we get

$$E[e^{it(X+Y)}] = E[e^{itX}e^{itY}] = E[e^{itX}]E[e^{itY}].$$

□

We now have the results needed to state the main theorem of this section.

Theorem 5.5 (Inversion Formula)**.** *Let X be a random variable with distribution μ, distribution function F, and a characteristic function ϕ. Then for every $a < b$ where F is continuous at a and b,*

$$\mu[a,b] + \frac{1}{2}\mu(\{a,b\}) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it}\phi(t)dt$$

Proof. We first fix $T < \infty$. Notice that

$$\Big|\frac{e^{-ita} - e^{-itb}}{it}\Big| = \Big|\int_a^b e^{-ity}dy\Big| \leq \int_a^b |e^{-ity}|dx = b - a.$$

This shows that the integral is indeed well defined as it is bounded. This allows for the application of Fubini's theorem.

$$\begin{aligned}
\frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it}\phi(t)dt &= \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it}\Big[\int_{-\infty}^{\infty} e^{-itx}\mu(dx)\Big]dt \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\Big[\int_{-T}^{T} \frac{e^{it(x-a)} - e^{it(x-b)}}{it}dt\Big]\mu(dx) \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\Big[\int_{-T}^{T} \frac{\sin(t(x-a)) - \sin(t(x-b))}{t}dt\Big]\mu(dx)
\end{aligned}$$

Since the cosine terms create an odd function, we can drop the integral and reach the last equality. By a change of variables $x = \frac{t}{c}$, we can obtain

$$\int_{-T}^{T} \frac{e^{icx}}{ix}dx = \int_{-T}^{T} \frac{\sin(cx)}{x}dx = 2\int_0^{|c|T} \frac{\sin(t)}{t}dt.$$

We should consider two important cases:

(1) If either $x < a$ or $b < x$,

$$\lim_{T\to\infty}\int_{-T}^{T} \frac{e^{it(x-a)} - e^{it(x-b)}}{it}dt = \lim_{T\to\infty}\Big[\int_{-T}^{T} \frac{\sin(t(x-a)) - \sin(t(x-b))}{t}dt\Big] = 0.$$

(2) If $a < x < b$,

$$\lim_{T\to\infty}\Big[\int_{-T}^{T} \frac{\sin(t(x-a)) - \sin(t(x-b))}{t}dt\Big] = 4\lim_{T\to\infty}\int_0^T \frac{\sin(t)}{t}dt = 2\pi,$$

which follows from the fact that as $T \to \infty$, $\int_0^T \frac{\sin(x)}{x}dx \to \frac{\pi}{2}$.

(3) If $x = a$ or $x = b$,

$$\lim_{T\to\infty}\Big[\int_{-T}^{T}\frac{\sin\left(t(x-a)\right)-\sin\left(t(x-b)\right)}{t}dt\Big]=\pi.$$

Since

$$g(a,b,x;T)=\int_{-T}^{T}\frac{e^{it(x-a)}-e^{it(x-b)}}{it}dt$$

is uniformly bounded in a, b, x, and T, then we can use the dominated convergence theorem to show that

$$\lim_{T\to\infty}\frac{1}{2\pi}\int_{-\infty}^{\infty}\Big[\int_{-T}^{T}\frac{e^{it(x-a)}-e^{it(x-b)}}{it}dt\Big]\mu(dx)=\mu(a,b)+\frac{1}{2}\mu(\{a,b\})$$

□

From the Inversion Formula, we can see that there exists a bijection between random variables and their respective characteristic functions. Therefore, if we can find the characteristic function of a random variable, the inversion formula will allow us to find the distribution function.

Proposition 5.6. *Suppose X is a random variable with the characteristic function ϕ where $E[|X|] < \infty$. Then ϕ is continuously differentiable and*

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

Proof. First, note that for all real c,

$$|e^{ic}-1| \le |c|.$$

We verify this by observing that

$$|e^{ic}-1| = \Big|\int_0^c ie^{ix}dx\Big| \le \int_0^c |ie^{ix}|dx = c.$$

We have the following:

$$\frac{\phi(t+\delta)-\phi(t)}{\delta}=\int_{-\infty}^{\infty}\frac{e^{i(t+\delta)x}-e^{itx}}{\delta}\mu(dx).$$

Note that

$$\Big|\frac{e^{i(t+\delta)x}-e^{itx}}{\delta}\Big| \le \frac{|\delta x|}{x} \le |x|.$$

Since $E[|X|]$ is finite, we know that $|x|$ is integrable with respect to the density μ. By the dominated convergence theorem and the limit definition of derivatives, we get

$$\lim_{\delta\to 0}\int_{-\infty}^{\infty}\frac{e^{i(t+\delta)x}-e^{itx}}{\delta}\mu(dx)=\int_{-\infty}^{\infty}ixe^{itx}\mu(dx)=\phi'(t).$$

Hence,

$$\phi'(0)=i\int_{-\infty}^{\infty}x\mu(dx)=iE[X]$$

□

We will now provide a generalization for random variables with finite higher-order moments.

Proposition 5.7. *Let X be a random variable with a characteristic function ϕ such that $E[|X|^k] < \infty$ for some positive integer k. Then ϕ has k continuous derivatives and*

$$\phi^{(j)}(t) = i^j E[X^j e^{itX}], j = 1, 2, 3, \dots, k$$

Proof. We will reach the conclusion with an induction on j. From proposition 5.6, we can see that the hypothesis holds true for the $j = 1$ case. We will assume it is true for some $j < k$.

$$\begin{aligned}\phi^{(j+1)}(t) &= \frac{d}{dt} i^j E[X^j e^{itX}] \\ &= i^j E[X^j (iXe^{itX})] \\ &= i^{(j+1)} E[X^{(j+1)} e^{itX}]\end{aligned}$$

By induction, the claim holds. □

6. Central Limit Theorem

Lemma 6.1. *Consider a random variable X with a density distribution by the standard normal distribution, $\mathcal{N}\{0, 1\}$. The characteristic function ϕ of X is*

$$\phi(t) = e^{-\frac{t^2}{2}}.$$

Proof. First, recall that

$$\phi(t) = E[e^{itX}] = \int_{-\infty}^{\infty} e^{itx} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

By some algebraic manipulation, we get

$$\phi(t) = \int_{-\infty}^{\infty} [\cos(tx) + i \sin(tx)] \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \int_{-\infty}^{\infty} [\cos(tx)] \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

since $i \sin(tx)$ is an odd function. By differentiating with respect to t, we get

$$\phi'(t) = \int_{-\infty}^{\infty} [-x \sin(tx)] \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \int_{-\infty}^{\infty} [-t \cos(tx)] \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

We can see that $\phi'(t) = -t\phi(t)$, which implies that $\frac{d}{dt}\phi(t)e^{\frac{t^2}{2}} = 0$ and $\phi(t)e^{\frac{t^2}{2}}$ is constant. As $\phi(0) = 1$, it follows that $\phi(t) = e^{-\frac{t^2}{2}}$. □

We are now ready to prove the classical central limit theorem.

Theorem 6.2 (Classical Central Limit Theorem)**.** *Let $X_1, X_2, \dots$ be independent, identically distributed random variables with finite mean μ and finite variance σ^2. If $S_n = X_1 + \dots + X_n$, then*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} = Z \to \mathcal{N}\{0, 1\}$$

where the distribution of Z converges in distribution to the standard normal as $n \to \infty$.

Proof. First, without loss of generality, consider the characteristic function ϕ_X for the random variable X with probability measure μ:

$$\phi_X(t) = E[e^{itX}] = \int_{-\infty}^{\infty} e^{itX} \mu(dx).$$

By Taylor's theorem, we can expand the exponential term following proposition 5.7:

$$e^{itX} = 1 + itX - \frac{t^2}{2}X^2 + o(t^2)$$

Here, $o(t^2)$ is an infinite polynomial in X where $\frac{o(t^2)}{t^2} = 0$ as $t \Rightarrow 0$. Applying this expansion, we get

$$\phi_X(t) = E[e^{itX}] = 1 + itE[X] - \frac{t^2}{2}E[X^2] + o(t^2).$$

Notice that $E[X]$ is the mean and $E[X^2]$ is the variance. In particular, if the random variable X has a mean of 0 and a variance of 1, then

$$\phi_X(t) = 1 - \frac{t^2}{2} + o(t^2)$$

Now, consider the sample mean $\overline{X_N}$ of N i.i.d random variables $X_i \sim (\mu, \sigma^2)$, where

$$\overline{X}_N = \frac{\sum_{i=1}^N X_i}{N}.$$

(i) $E[\overline{X}_N] = \frac{1}{N}\sum_{i=1}^N E[X_i] = \frac{1}{N}N\mu = \mu$
This follows from the properties of expected value and that $E[X_i] = \mu$ for all i.

(ii) $E[(\overline{X}_N)^2] = Var[(\overline{X}_N)] = \frac{1}{N^2}\sum_{i=1}^N Var[X_i] = \frac{1}{N^2}(N\sigma^2) = \frac{\sigma^2}{N}$
This follows from the fact that each X_i is i.i.d. and is thus orthogonal.

We will now standardize $\overline{X}_N$ to obtain a random variable with mean of 0 and variance of 1. Consider the following random variable,

$$Z_N = \frac{N\overline{X}_N - N\mu}{\sigma\sqrt{N}} = \frac{\sum_{i=1}^N (X_i - \mu)}{\sigma\sqrt{N}} = \frac{\sum_{i=1}^N Y_i}{\sqrt{N}}$$

where $Y_i = \frac{X_i - \mu}{\sigma}$. We can see that $E[Y_i] = 0$ and $Var[Y_i] = \frac{Var[X_i]}{\sigma^2} = 1$. Then for Y_i, the characteristic function can be expressed by

$$\phi_Y = 1 - \frac{t^2}{2} + o(t^2).$$

By defining $Z_N = \sum_{i=1}^N \frac{Y_i}{\sqrt{N}}$, we can then find the characteristic function of Z_N by noting that each Y_i is i.i.d. We then have the following:

$$\phi_{Z_N}(t) = [\phi_Y(\frac{t}{\sqrt{N}})]^N = [1 - \frac{t^2}{2N} + o(\frac{t^2}{N})]^N.$$

Claim: for all real t, $\phi_{Z_N}(t)$ converges pointwise to $e^{\frac{-t^2}{2}}$ as N approaches ∞. We know that for a sequence $\{a_n\} \to a \in \mathbb{R}, \lim_{n\to\infty} \frac{a_n}{n} \to 0$. From the limit definition of derivatives, we have

$$\lim_{n\to\infty} \frac{\ln(1 + \frac{a_n}{n})}{\frac{a_n}{n}} = \frac{d}{dx}\ln(x)\bigg|_{x=1} = 1.$$

Therefore,

$$\lim_{N\to\infty} \frac{\ln[1 - \frac{t^2}{2N} + o(\frac{t^2}{N})]}{-\frac{t^2}{2N} + o(\frac{t^2}{N})} = 1.$$

By rearranging terms, we get

$$\lim_{N\to\infty} N\ln[1-\frac{t^2}{2N}+o(\frac{t^2}{N})] = -\frac{t^2}{2}.$$

This ultimately shows that

$$\lim_{N\to\infty}[1-\frac{t^2}{2N}+o(\frac{t^2}{N})]^N = e^{\frac{-t^2}{2}},$$

which is the desired characteristic function of the standard normal distribution. □

Acknowledgments

First, I would like to thank my mentor, Mariya Sardarli, for supporting my mathematical interests and for guiding me through the research process. Without her guidance, this paper would not have been possible. I would also like to thank Daniil Rudenko for hosting the apprentice lectures. Finally, I want to thank Peter May for hosting the REU program.

Bibliography

References

[1] Lawler, Greg. "Notes on Probability." The University of Chicago, Chicago, IL, March 21, 2016.
[2] Durrett, Rick. Probability: Theory and Examples. 5th ed. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press, 2019.
[3] Stein, Elias M., and Rami Shakarchi. Real Analysis: Measure Theory, Integration, and Hilbert Spaces. Princeton; Oxford: Princeton University Press, 2005.
[4] Hunter, John K. "Measure Theory." University Lecture Notes, Department of Mathematics, University of California at Davis.