PROBABILITY THEORY, FOURIER TRANSFORM AND CENTRAL LIMIT THEOREM

by

YEGOR SOROKIN

B.S., Dnipropetrovsk National University, Ukraine, 2006

A REPORT

submitted in partial fulfillment of the
requirements for the degree

MASTER OF SCIENCE

Department of Mathematics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas
2009

Approved by:

Major Professor
David Auckly

Abstract

In this report we present the main concepts of probability theory: sample spaces, events, random variables, distributions, independence, central limit theorem. Most of the material may be found in the notes of Bass[1]. The work is motivated by wide range of applications of probability theory in quantitative finance.

Table of Contents

List of Figures

Chapter 1

Basic notions

Random experiments are experiments whose output cannot be surely predicted in advance. We are interested in different outcomes of an experiment, which we call events. The set of all possible states of an experiment is a sample space and is usually denoted by Ω.

As an example of a random experiment consider a roll of a fair six-sided die. There are 6 possible outcomes (states): after the throw, a die's facet has 1, 2, 3, 4, 5 or 6. So for this particular case the sample space is

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

By putting together the sample space we have described all possible outcomes such as "we get 5". Surprisingly enough, element $5 \in \Omega$ represents this case. On the other hand, we may have a more complex statement like "we do not get 5" on the facet of the die. Please note that there is no single element from the sample space that would describe this statement. We need a subset of the sample space instead: $\{1, 2, 3, 4, 6\} \subseteq \Omega$ encodes the "we do not get 5" statement. Also we may think of the following statement: "we get 3 or 4", which is a union of two elements of the sample space. Not only do the elements of the sample space reflect the statements about possible outcomes of the process, but also the subsets of the sample space may have a meaning. So we also want to encode the complements of the outcomes (like in the case with element 5 above) as well as unions of all available subsets of the sample space. In order to capture all possible statements we introduce the notion of the σ-algebra.

Definition 1.1. *A subset $\mathcal{F}$ of the power set of a set Ω is a σ-algebra if it has the following properties:*

1. *$\mathcal{F}$ contains the set Ω as an element.*
2. *If E is in $\mathcal{F}$ then so is the complement $(X \setminus E)$ of E.*
3. *The union of countably many sets in $\mathcal{F}$ is also in $\mathcal{F}$.*

Examples of σ-algebras are given by full power set of Ω and the family consisting only of an empty set and set Ω itself. Let's take a look at a more interesting example of σ-algebra. In the above example with a fair 6-sided die, set set $X = \{1, 2, 3, 4, 5, 6\}$ and σ-algebra

$$\mathcal{F} = \{\emptyset, \{1, 2, 3\}, \{4, 5, 6\}, \{1, 2, 3, 4, 5, 6\}\}$$

Another example of σ-algebra is the family of all Borel sets of $\mathbb{R}$. This is the smallest σ-algebra containing the open sets. In the finite/countable cases we usually take the full power set of Ω as a σ-algebra, while in the uncountable case usually take a Borel σ-algebra.

While the outcome (state) is represented by only 1 element of the sample space, an event is an element of the σ-algebra i.e. a subset of sample space. A state describes the result of the experiment, while event contains interpretation of the desired outcome. For example if one wins if the number on the die's facet is even, one may want the die to show even number after the throw, and this is a subset of a sample space rather than a single element: outcomes 2,4,6 all are even numbers.

Definition 1.2. *An event is an element of the σ-algebra.*

For the purpose of illustrating the idea of an event consider the following example: imagine that we have 3 fair coins, we flip each coin and observe the final combination of heads and tails on the table. For each coin we may denote H as getting a head and T as getting a tail. We may encode the final triples for this experiment that make the sample space using the H/T convention:

$$\Omega = \{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}$$

Let's say we are interested in the total number of heads after each of 3 coins landed on the table. Keeping this goal in mind, we define events that are of interest for us in a following manner: "we get 3 heads" corresponds to $\{HHH\} \in \mathcal{F}$, "we get 2 heads" corresponds to $\{HHT, HTH, THH\} \in \mathcal{F}$, etc.

One may want to get a feeling of the process and/or have a way of predicting, measuring the process. Thus we assign a numerical value that describes the likelihood of occurrence to each event - its probability. In the above example with a fair die, intuitively the chance of getting one of 6 facets should be the same.

However processes in the world are not limited to finite number of possible outcomes. For example the lifetime of the light bulb measured in seconds may have practically infinite amount of possibilities. Now defining an event as well as assigning a numerical value that describes likelihood of the event becomes somewhat harder. We will need a notion of measure in order to define probabilities of events in uncountable sample space. And we rely on the definition of σ-algebra to define measure.

Definition 1.3. *A measure μ is a function defined on a σ-algebra $\mathcal{F}$ over a set Ω and taking values in the extended interval $[0; \infty]$ such that the following properties are satisfied:*

- *The empty set has measure zero: $\mu(\emptyset) = 0$.*

- *If $E_1, E_2, E_3, \ldots$ is a countable sequence of pairwise disjoint sets in Σ, the measure of the union of all the E_i is equal to the sum of the measures of each E_i:*

$$\mu(\bigcup_{k=1}^{r} E_i) = \sum_{k=1}^{n} \mu(E_i)$$

The simplest example of a measure is the counting measure: $\mu(S) =$ number of elements in a set S.

Definition 1.4. *A probability measure $\mathbb{P}$ is a measure with total mass* 1 *i.e.* $\mathbb{P}(\Omega) = 1$.

Probability measures give a way of assigning probabilities to events. For the example with 3 coin tosses we may construct a probability measure in the following way:

$$\mathbb{P} = \frac{\#\text{ of favourable states in the event}}{\text{total }\#\text{ of states}}$$

Consider for instance the following event: one gets 3 heads. Then in the language of the σ-algebras it is $\{HHH\} \subset \mathcal{F}$. Please note that there is 1 state in the event, while the total number of states is 8. According to the measure defined above the probability of the event is:

$$\mathbb{P} = \frac{\#\text{ of states in the event}}{\text{total }\#\text{ of states}} = \frac{1}{8}$$

Definition 1.5. *A probability space is a triple* $(\Omega, \mathcal{F}, \mathbb{P})$*, where* Ω *is a set called sample space,* $\mathcal{F}$ *is a* σ*-algebra and* $\mathbb{P}$ *is a probability measure.*

Definition 1.6. *A measurable space (or* σ*-field) is a pair* $(\Omega, \mathcal{F})$*, where* Ω *is a sample space and* $\mathcal{F}$ *is a* σ*-algebra on* Ω.

Let's say we have a dart board with 3 sections (Figure (1.1)): a circle with a relatively small radius in the center ("Bull's eye", red), big concentric circle that touches the edges of the square dart board ("outer bull", green) and 4 remaining areas at the corners of the board ("Milk", white). We assume that the person who throws the darts will inevitably hit the board and cannot miss the board. In other words the darts will end up either in "Bull's eye" or in circled area or corners of the board. Let's say the radius of the "Bull's eye" is 1, the radius of the big circle is 5, thus the side of the square has size 10. Once we have the dimensions of the figures, we may compute the probabilities of hitting certain area on the board. One more assumption: the board is "fair" - hitting any point in any area of the board has equal likelihood. We may also say that there are no cheating in the game: the back of the board does not contain magnets to attract the dart that may contain metal parts. With this assumptions in mind one may find the probabilities of hitting certain area of the board by looking at the ratio of the region's area to the area of the whole board: "Bull's eye" - $\frac{\pi}{100}$; "outer bull" - $\frac{24\pi}{100}$; "Milk" - $\frac{100-25\pi}{100}$.

In the example with tossing a coin since each outcome uniquely determines the number of heads in each trial, we may think of it as a function. Let C be the number of heads, then

$$C(HHH) = 2, C(TTH) = 1, C(THT) = 1, \ldots, C(TTT) = 0$$

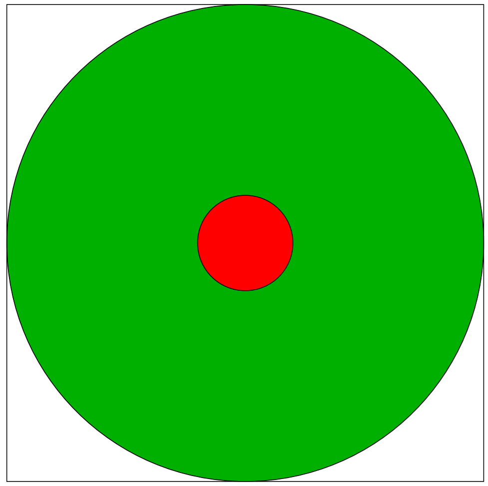

Figure 1.1: *Dartboard*

Definition 1.7. *The inverse image of a set A under a function f is $f^{-1}(A) = \{x \in X | f(x) \in A\}$.*

Definition 1.8. *Let Σ and T be σ-algebras over sets X and Y correspondingly. Then $f : X \to Y$ is measurable if every $A \in T$, $f^{-1}(A) \subset \Sigma$.*

Definition 1.9. *A measurable function X from Ω to $\mathbb{R}$ is called a random variable.*

The head counting function C is an example of a random variable. We also may create a random variable for the example with the darts game above. Assume that hitting a certain area gives the player a number of points: 20 for the "Bull's eye" (denoted by B), 5 for the "outer bull" (denoted by O) and 0 for the "Milk" (denoted by M). For this example, the sample space is the square $[-5, 5]^2$. The σ-algebra of events is the algebra of Borel sets and the measure is unit normalized area measure.

Let S be the number of points earned for hitting the board after 1 throw. Then for example $S = 20\chi_B + 5\chi_O$. So S is a random variable.

Definition 1.10. *If a sample space Ω is countable, then random variable defined on it is called discrete.*

The random variable C is discrete.

Definition 1.11. *If a sample space Ω is not countable, then random variable defined on it is called continuous.*

The random variable S is continuous. The normal distribution is the most important example of a continuous probability space, however we will delay the description of it.

In order to proceed we will need the following lemma:

Lemma 1.12. *Let X, f be functions. Then $(f \circ X)^{-1}(A) = X^{-1}(f^{-1}(A))$.*

Proof. Let X, f be as above. Choose $y \in (f \circ X)^{-1}(A)$. Then $(f \circ X)(y) \in A$ by definition of inverse image. Therefore $f(X(y)) \in A$ by definition of the composition of functions. So $X(y) \in f^{-1}(A)$ (definition of inverse image) and $y \in X^{-1}(f^{-1}(A))$ Now choose $y \in X^{-1}(f^{-1}(A))$. Then similarly $y \in (f \circ X)^{-1}(A)$. So $(f \circ X)^{-1}(A) = X^{-1}(f^{-1}(A))$ as sets. □

Let's prove a simple result about random variables.

Theorem 1.13. *Let X be a random variable, f a Borel measurable function, then f(X) is a random variable.*

Proof. Let X, f be as above. Choose $A \subset \mathbb{R}$ Borel measurable. Now $(f \circ X)^{-1}(A) = X^{-1}(f^{-1}(A))$. Since $f^{-1}(A) \subset \mathcal{F}$ by definition of Borel measurable, and X is random variable itself, $X^{-1}(f^{-1}(A)) \subset \mathcal{F}$. So f(X) is a random variable by its definition. □

Once we defined a sample space and have a random variable, we want to get a feeling of where the outcomes will gravitate with time. For example, if I buy a lottery ticket, I should not expect to win a million Canadian dollars right away.

Definition 1.14. *The expectation of a random variable X is the integral of X with respect to* $\mathbb{P}$*:* $\mathbb{E}[X] = \int X(\omega)\mathbb{P}(d\omega)$

For the discrete random variables the expectation is just the probability-weighted sum of the possible values. In the example with coin toss :

$$\mathbb{E}[C] = 3 \cdot \frac{1}{8} + 2 \cdot \frac{3}{8} + 1 \cdot \frac{3}{8} + 0 \cdot \frac{1}{8} = 1.5$$

The expectation tells us that with growing number of experiments the average number of the heads after 3 coin tosses will be 1.5, which pretty much agrees with our intuition. Please note that expected value should not necessary be the element of the sample space.

Definition 1.15. *If X is a random variable and has expectation* $\mathbb{E}(X) = \mu$*, then the variance Var(X) of X is given by* $Var(X) = \mathbb{E}[(X - \mu)^2]$.

For a discrete random variable X with $\mathbb{E}(X) = \mu$, the variance is $Var(X) = \sum_{i=1}^{n} \mathbb{P}(\omega_i)(X(\omega_i) - \mu)^2$. Let's exploit the example with 3 consecutive coin tosses once again:

$$Var(C) = \frac{1}{8} \cdot (3 - 1.5)^2 + \frac{3}{8} \cdot (2 - 1.5)^2 + \frac{3}{8} \cdot (1 - 1.5)^2 + \frac{1}{8} \cdot (0 - 1.5)^2 = .75$$

One may think of a random variable as a way of partitioning the sample space: from all the variety of possible subsets we choose only limited amount. In the coin example we end up with 4 subsets:

$$\{HHH\}, \{HHT, HTH, THH\}, \{HTT, THT, TTH\}, \{TTT\}$$

Here the criterion for partitioning was the number of heads one observes at the end of the experiment. We already have probabilities for the separate outcomes, however with random variable not only does one deal with just states $\{HHH\}, \{TTT\}$, we also have subsets of the sample space $\{HHT, HTH, THH\}, \{HTT, THT, TTH\}$. So we want to use the knowledge of the probabilities of states to assign/compute probabilities of events.

We start with sample space and probabilities of the separate states, then we introduce events as a subsets of the sample space. At the same time from all the variety of available subsets we are interested only in certain number of subsets, and we would like to know the probabilities of events occurring.

Definition 1.16. *The law or distribution of X is the probability measure $\mathbb{P}_X$ on $\mathbb{R}$, given by*

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in A\}) = \mathbb{P}(X^{-1}(A))$$

Put differently while the probability measure assigns probabilities to the events (subsets of the sample space), the distribution of the random variable assesses the probability that the random variable takes values in set A.

In the example with 3 consecutive coin tosses let's describe the distribution of C. We will have following picture that corresponds to 4 possible outcomes:

$$\mathbb{P}(C^{-1}(\{0\})) = \mathbb{P}(\{\omega \in \Omega | C(\omega) \in \{0\}\}) = \mathbb{P}(\{TTT\}) = \frac{1}{8}$$
$$\mathbb{P}(C^{-1}(\{1\})) = \mathbb{P}(\{\omega \in \Omega | C(\omega) \in \{1\}\}) = \mathbb{P}(\{HTT, THT, TTH\}) = \frac{3}{8}$$
$$\mathbb{P}(C^{-1}(\{2\})) = \mathbb{P}(\{\omega \in \Omega | C(\omega) \in \{2\}\}) = \mathbb{P}(\{HHT, THH, HTH\}) = \frac{3}{8}$$
$$\mathbb{P}(C^{-1}(\{3\})) = \mathbb{P}(\{\omega \in \Omega | C(\omega) \in \{3\}\}) = \mathbb{P}(\{HHH\}) = \frac{1}{8}$$

We worked out cases when the value of the random variable C was feasible: after 3 coin tosses one may get either 0, 1, 2 or 3 heads. But these are not the only values on $\mathbb{R}$ that one may use. Let's explore other possibilities. What's the probability that one gets 5 heads after 3 coin tosses?

$$\mathbb{P}(C^{-1}(\{5\})) = \mathbb{P}(\{\omega \in \Omega | C(\omega) \in \{5\}\}) = \mathbb{P}(\emptyset) = 0$$

since $\mathbb{P}(\emptyset) = 0$ by the axioms of probability. This result comes as no surprise: one cannot get 5 heads after 3 coin tosses, even if heads appear after each coin toss. Similarly

$$\mathbb{P}(C^{-1}(\{-2\})) = \mathbb{P}(\{\omega \in \Omega | C(\omega) \in \{-2\}\}) = \mathbb{P}(\emptyset) = 0$$
$$\mathbb{P}(C^{-1}((-\infty, -1])) = \mathbb{P}(\{\omega \in \Omega | C(\omega) \in (-\infty, -1]\}) = \mathbb{P}(\emptyset) = 0$$

We now consider the normal distribution and normally distributed random variables.

Definition 1.17. *The Gaussian (also called normal) distribution is given by the probability measure on $\mathbb{R}$ with the Borel σ-algebra*

$$\mathbb{P}(X \in A) = \int_A \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx, A \text{ Borel}$$

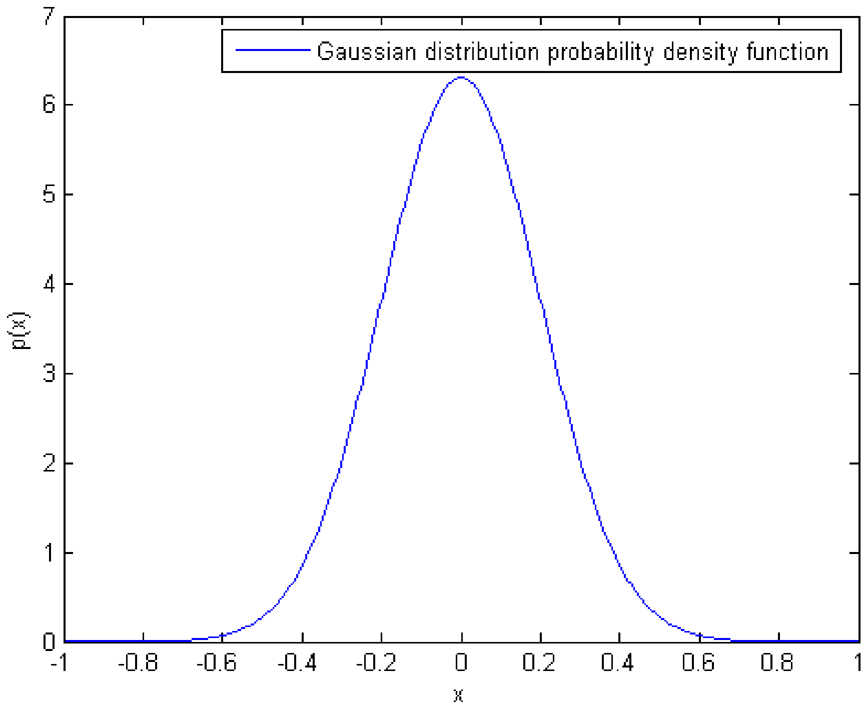

Figure 1.2: *Gaussian distribution probability density function*

The graph of the density is illustrated on the Figure (1.2).

Definition 1.18. *The cumulative distribution function of a random variable X is the function $F_X(t) = \mathbb{P}_X((-\infty, t])$.*

Consider the cumulative distribution function for the random variable C (number of heads after 3 coin flips).

$$F_C(0) = \mathbb{P}_C((-\infty, 0]) = \mathbb{P}(\{TTT\}) = \frac{1}{8}$$

Similarly

$$F_C(1) = \mathbb{P}_C((-\infty, 1]) = \mathbb{P}(\{TTT, HTT, THT, TTH\}) = \frac{4}{8} = \frac{1}{2}$$

$$F_C(2) = \mathbb{P}_C((-\infty, 2]) = \mathbb{P}(\{TTT, HTT, THT, TTH, HHT, HTH, THH\}) = \frac{7}{8}$$

$$F_C(3) = \mathbb{P}_C((-\infty, 3]) = \mathbb{P}(\{TTT, HTT, THT, TTH, HHT, HTH, THH, HHH\}) = \frac{8}{8} = 1$$

It's easy to check that for $x \in (-\infty, 0), F_C(x) = 0; x \in [0, 1), F_C(x) = \frac{1}{8}; x \in [1, 2), F_C(x) = \frac{1}{2}; x \in [2, 3), F_C(x) = \frac{7}{8}; x \in [3, +\infty), F_C(x) = 1$. One may find a graph of the cumulative distribution function of C at the Figure (1.3).

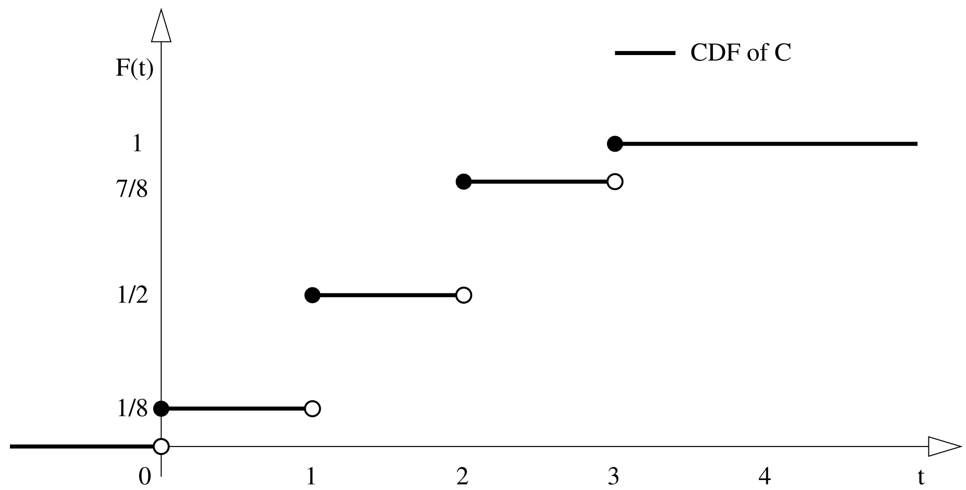

Figure 1.3: *Cumulative distribution function of C*

For Gaussian random variable the cumulative distribution function is given by

$$F_G(x) = \frac{1}{2}(1 + erf(\frac{x}{\sqrt{2}}))$$

(see Figure (1.4)).

It is interesting to ask what functions can be cumulative distribution functions. The first observation is that a cumulative distribution function is necessarily increasing:

Lemma 1.19. *If $s \leq t$, then $F_X(s) \leq F_X(t)$.*

Proof. Choose s,t such that $s \leq t$. Then $F_X(s) = \mathbb{P}_X((-\infty, s])$. $F_X(t) = \mathbb{P}((-\infty, t]) = \mathbb{P}((-\infty, s] \cup [s, t]) = \mathbb{P}((-\infty, s]) + \mathbb{P}((s, t])$. And since $\mathbb{P}_X((s, t]) \geq 0$, $F_X(s) \leq F_X(t)$. $\square$

Now consider the following function:

$$F(t) = \begin{cases} 0 & \text{if } t < 0 \\ \frac{1}{2} & \text{if } t = 0 \\ 1 & \text{if } t > 0 \end{cases}$$

For any $\epsilon > 0, \mathbb{P}(X \in (-\infty, \epsilon]) = 1$.

We see that

$$\mathbb{P}((-\infty, 1]) = F(1) = 1$$
$$\mathbb{P}((-\infty, \frac{1}{2}]) = F(\frac{1}{2}) = 1$$

Since $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$, then $\mathbb{P}(B) = \mathbb{P}(A \cup B) - \mathbb{P}(A)$. We see $\mathbb{P}((\frac{1}{2}, 1]) = \mathbb{P}((-\infty, \frac{1}{2}] \cup (\frac{1}{2}, 1]) - \mathbb{P}((-\infty, \frac{1}{2}]) = \mathbb{P}((-\infty, 1]) - \mathbb{P}((-\infty, \frac{1}{2}]) = 1 - 1 = 0$. Similarly $\mathbb{P}((\frac{1}{4}, \frac{1}{2}]) = 0$, etc. Thus $\mathbb{P}((0, 1]) = \mathbb{P}((\frac{1}{2}, 1]) + \mathbb{P}((\frac{1}{4}, \frac{1}{2}]) + \ldots = 0 + 0 + \ldots = 0$, implying that $1 = \mathbb{P}((-\infty, 1]) = \mathbb{P}((-\infty, 0]) + \mathbb{P}((0, 1]) = F(0) + 0 = \frac{1}{2} + 0 = \frac{1}{2}$, and this cannot happen.

In fact we see that cumulative distribution function is right continuous:

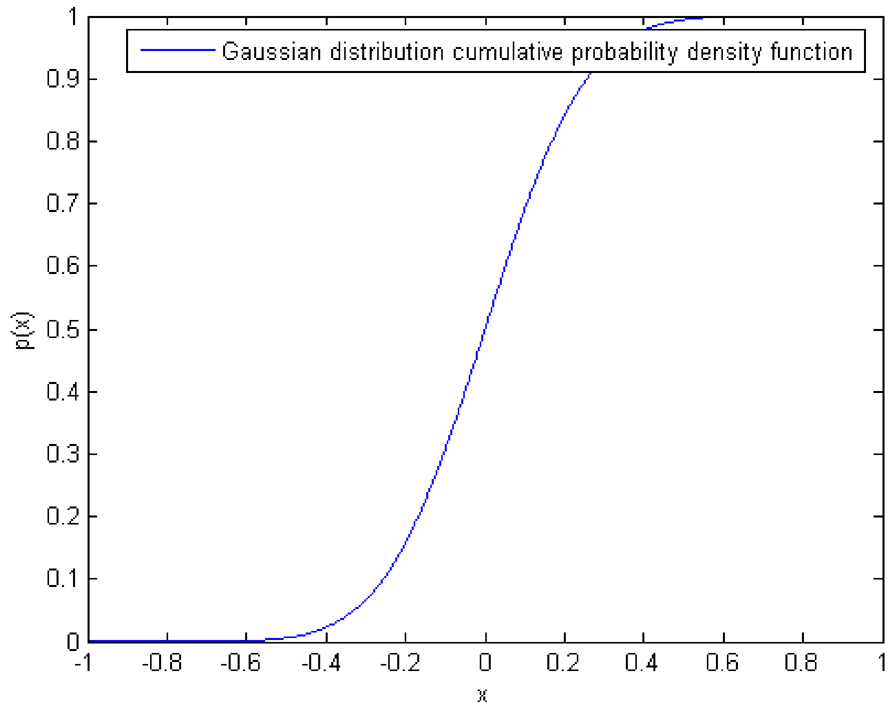

Figure 1.4: *Gaussian distribution cumulative distribution function*

Lemma 1.20. *The cumulative distribution function is right continuous.*

Proof. Let x_n decrease to x. Then $\bigcap_{n=1}^{\infty}(-\infty, x_n] = (-\infty, x]$ and the sequence of events $\{(-\infty, x_n]\}$ is a decreasing sequence. Consider $\mathbb{P}((-\infty, x + x_n]) - \mathbb{P}((-\infty, x]) = \mathbb{P}((x, x + x_n]) = \int \chi_{(x,x+x_n]}(t)d\mathbb{P}(t)$. Now $\lim_{n\to+\infty}\int \chi_{(x,x+x_n]}(t)d\mathbb{P}(t) = \int \lim_{n\to+\infty} \chi_{(x,x+x_n]}(t)d\mathbb{P}(t) = 0$ by the Lebesgue dominated convergence theorem. Thus $\lim_{n\to+\infty}(\mathbb{P}((-\infty, x + x_n]) - \mathbb{P}((-\infty, x])) = 0$, in other words $\lim_{n\to+\infty}\mathbb{P}((-\infty, x + x_n]) = \mathbb{P}((-\infty, x])$. $\square$

Lemma 1.21. $\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$

Proof. $1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup A^C) = \mathbb{P}(A) + \mathbb{P}(A^C)$, since A, A^C are disjoint. So $\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$ by algebra. $\square$

Lemma 1.22. *Define $\mathcal{B} := \cap\{\mathcal{F}|\mathcal{F}$ is a σ-algebra and (a,b) $\in \mathcal{F}$ for all (a,b)$\}$. Then $\mathcal{B}$ is a σ-algebra.*

Proof. Choose any $A \in \mathcal{B}$. Then $A \in \mathcal{F}$, so $A^c \in \mathcal{F}$ since $\mathcal{F}$ is a σ-algebra, and so $A^c \in \mathcal{B}$ by definition of $\cap$. Choose a set $\{A_n\} \in \mathcal{B}$. Then $A_n \in \mathcal{F}$, so $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$, since $\mathcal{F}$ is a σ-algebra, and so $\bigcup_{n=1}^{\infty} A_n \in \mathcal{B}$ by definition of $\cap$. Obviously $\mathbb{R} \in \mathcal{B}$. Thus $\mathcal{B}$ is a σ-algebra by definition of σ-algebra. $\square$

We will use the following fact in the next lemma:

Fact 1.23. *A Borel set is a set that is obtained by repeatedly taking countable unions and complements of open sets.*

Let us denote the result of N complements and/or countable unions of open sets as a "level N Borel set".

Lemma 1.24. *If $F_X(t) = F_Y(t)$, then $\mathbb{P}_X = \mathbb{P}_Y$.*

Proof. Proceed by induction. Base case (level 0 Borel set i.e. (a, b)): $\mathbb{P}_X((a, \infty)) = \mathbb{P}_X((-\infty, a]^c) = 1 - \mathbb{P}_X((-\infty, a]) = 1 - F_X(a) = 1 - F_Y(b) = \mathbb{P}_Y((-\infty, a)) = \mathbb{P}_Y((-\infty, a]^c) = \mathbb{P}_Y((a, \infty))$ and $\mathbb{P}_X((a, b]) = \mathbb{P}_X((-\infty, b] \cap (a, \infty)) = \mathbb{P}_X((-\infty, b]^{cc} \cap (a, \infty)^{cc}) = \mathbb{P}_X(((-\infty, b]^c \cup (a, \infty)^c)^c) = 1 - \mathbb{P}_X((-\infty, b]^c \cup (a, \infty)^c) = 1 - \mathbb{P}_X((-\infty, b]^c) - \mathbb{P}_X((a, \infty)^c) = 1 - (1 - \mathbb{P}_X((-\infty, b])) - \mathbb{P}_X((-\infty, a]) = F_X((-\infty, b])) - F_X((-\infty, a]) = F_Y((-\infty, b])) - F_Y((-\infty, a])$ and repeat the argument backwards with Y. Now let $b_n \nearrow b$, $b_0 = a, b_k \geq a$. So

$$
\begin{aligned}
\mathbb{P}_X((a, b)) &= \mathbb{P}_X(\bigcup_{k=1}^{\infty}(b_k, b_{k+1}]) = \\
&= \prod_{k=1}^{\infty} \mathbb{P}_X(b_k, b_{k+1}]) = \\
&= \prod_{k=1}^{\infty} \mathbb{P}_Y(b_k, b_{k+1}]) = \\
&= \mathbb{P}_Y(\bigcup_{k=1}^{\infty}(b_k, b_{k+1}]) = \\
&= \mathbb{P}_Y((a, b))
\end{aligned}
$$

Induction step: assume $\mathbb{P}_X(B) = \mathbb{P}_Y(B)$ for all level N Borel sets as our induction hypothesis. Let A be a level N Borel set, then either $A = B^c$, or $A = \bigcup_{k=1}^{\infty} B_k$.

- case 1: $A = B^c$. Then $\mathbb{P}_X(A) = \mathbb{P}_X(B^c) = 1 - \mathbb{P}_X(B) = 1 - \mathbb{P}_Y(B) = \mathbb{P}_Y(B^c) = \mathbb{P}_Y(A)$.
- case 2: $A = \bigcup_{k=1}^{\infty} B_k$. Rewrite expression for A to make sure that sets are disjoint: $A = \bigcup_{j=1}^{\infty}(B_{j+1} - \bigcup_{k=0}^{j} B_k)$. Then

$$
\begin{aligned}
\mathbb{P}_X(A) =& \mathbb{P}_X(\bigcup_{j=1}^{\infty}(B_{j+1} - \bigcup_{k=0}^{j} B_k)) = \sum_{j=1}^{\infty} \mathbb{P}_X(B_{j+1} - \bigcup_{k=0}^{j} B_k) = \\
=& \sum_{j=1}^{\infty} \mathbb{P}_Y(B_{j+1} - \bigcup_{k=0}^{j} B_k) = \mathbb{P}_Y(\bigcup_{j=1}^{\infty}(B_{j+1} - \bigcup_{k=0}^{j} B_k)) = \mathbb{P}_Y(A)
\end{aligned}
$$

and that completes the proof. □

Theorem 1.25. *Given a measure $\mathbb{P}$ on $\mathbb{R}$ with total mass 1 (probability measure), it is possible to construct a random variable X on $[0, 1]$ such that $\mathbb{P}_X = \mathbb{P}$.*

Proof. Let $\mathbb{P}$ be as above. Set the sample space $\Omega = [0,1]$ with the σ-algebra of the Borel sets and Lebesgue measure μ. Set $F_{\mathbb{P}}(t) = \int \chi_{(-\infty,t]}(s) d\mathbb{P}(s)$. $F_{\mathbb{P}}$ is monotone and continuous from the right by Lemma (1.20). Set $X : [0,1] \to \mathbb{R}$ to be

$$X(s) = \begin{cases} \sup\{t | F_{\mathbb{P}}(t) \leq s\} & \text{if } s \neq 1 \\ 0 & \text{if } s = 1 \end{cases}$$

If $s < 1$ and $a_N \to \infty$, then $\lim_{t\to\infty} F_{\mathbb{P}}(t) = \lim_{N\to\infty} F_{\mathbb{P}}(a_N) = \lim_{N\to\infty}(\int \chi_{(-\infty,a_N]}(s) d\,\mathbb{P}(s)) = \int \lim_{N\to\infty} \chi_{(-\infty,a_N]}(s) d\,\mathbb{P}(s) = \int 1 d\,\mathbb{P}(r) = 1$. So there is a t_s such that $F_{\mathbb{P}}(t_s) > s$. But $F_{\mathbb{P}}(t)$ is monotone, thus for $t > t_s, F_{\mathbb{P}}(t) \geq F_{\mathbb{P}}(t_s) > s$. In other words t_s is upper bound, therefore X is well-defined. We will show $[0, F_{\mathbb{P}}(t)) \subseteq X^{-1}((-\infty, t]) \subseteq [0, F_{\mathbb{P}}(t)]$, so X is measurable. To see the first set containment, let $a < F_{\mathbb{P}}(t)$, so $a \in [0, F_{\mathbb{P}}(t))$. Note that $F_{\mathbb{P}}(t) \leq 1$, so $a < 1$, and therefore $X(a) = \sup\{r | F_{\mathbb{P}}(r) \leq a\}$. Assume $X(a) > t$, then there exists $r > t$ such that $F_{\mathbb{P}}(r) \leq a$ by definition of sup. Finally using the initial assumption we have $F_{\mathbb{P}}(r) < F_{\mathbb{P}}(t)$. On the other hand F is monotone increasing: for $r > t$, $F_{\mathbb{P}}(r) \geq F_{\mathbb{P}}(t)$. So we have a contradiction. Now let us turn our attention to the proof of the second containment, namely $X^{-1}((-\infty, t]) \subseteq [0, F_{\mathbb{P}}(t)]$: let $a \in X^{-1}((-\infty, t])$, so $X(a) \in (-\infty, t]$ and $X(a) \leq t$. Assume $F_{\mathbb{P}}(t) < a$. Take $\epsilon = a - F_{\mathbb{P}}(t)$. Since $F_{\mathbb{P}}(t)$ is right continuous, there exists $\delta > 0$ such that for $r \in (t, t+\delta) \Rightarrow |F_{\mathbb{P}}(r) - F_{\mathbb{P}}(t)| < \epsilon$. Now $F_{\mathbb{P}}(r) < \epsilon + F_{\mathbb{P}}(t) = a - F_{\mathbb{P}}(t) + F_{\mathbb{P}}(t) = a$. At the same time $X(a) \geq r$ by definition of $X(a)$ and $r > t$ by continuity, which gives is a contradiction (initially we let $X(a) \leq t$), so $a \leq F_{\mathbb{P}}(t)$ and consequently $a \in [0, F_{\mathbb{P}}(t)] \Rightarrow X^{-1}((-\infty, t]) \subseteq [0, F_{\mathbb{P}}(t)]$. Now $F_{\mathbb{P}_X}(t) := \mu(X^{-1}((-\infty, t])) = F_{\mathbb{P}}(t)$, therefore $\mathbb{P}_{\mathbb{X}} = \mathbb{P}$. $\square$

Chapter 2

Independence

Intuitively events from unrelated experiments cannot influence the likelihood of occurrence of each other. A flip of a coin does not affect the result of trowing a die. Getting heads will not change the likelihood of getting 5 after a throw of a die. In a similar manner outcomes of 2 consecutive throws of a die are not related: if we get 2 after the first roll, this fact has no effect on seeing 4 after the second roll. On the other hand 2 experiments may be explicitly interconnected: let's say we expect the sum after two rolls of a die to be ten. Here getting a four the first time means that we would need to get an six on the second roll.

Definition 2.1. *Events A and B are independent if* $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. *More generally* $A_1, \ldots, A_n$ *are independent if for any subset* $\{i_1, \ldots, i_j\}$ *of* $\{1, \ldots, n\}$

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_j}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \cdots \mathbb{P}(A_{i_j})$$

Example of independent events: let's say we flip a coin and then roll a die.

$$\Omega = \{(H,1),(H,2),(H,3),(H,4),(H,5),(H,6),(T,1),(T,2),(T,3),(T,4),(T,5),(T,6)\}$$

Let event A be the following: "coin shows tails" event B - "die shows an even number". Then $A = \{(T,1),(T,2),(T,3),(T,4),(T,5),(T,6)\}$, $B = \{(H,2),(H,4),(H,6),(T,2),(T,4),(T,6)\}$, $A \cap B = \{(T,2),(T,4),(T,6)\}$ and $\mathbb{P}(A \cap B) = \frac{3}{12} = \frac{1}{4}$, while $\mathbb{P}(A) \cdot \mathbb{P}(B) = \frac{6}{12} \cdot \frac{6}{12} = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. So A and B are independent since $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Example of dependent events: let's say we roll a fair 6-sided die two times in a row. Let "get a four on the first roll" is event A, "the sum after two rolls is ten" is event B. Then $\Omega = \{(a,b)|a,b \in \{1,2,3,4,5,6\}\}$, $A = \{(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)\}$, $B = \{(4,6),(6,4),(5,5)\}$, $A \cap B = \{(4,6)\}$, and $\mathbb{P}(A \cap B) = \frac{1}{36}$, while $\mathbb{P}(A) \cdot \mathbb{P}(B) = \frac{6}{36} \cdot \frac{3}{36} = \frac{1}{72}$. Thus A and B are dependent since $\mathbb{P}(A \cap B) \neq \mathbb{P}(A)\mathbb{P}(B)$.

Lemma 2.2. *If A and B are independent events, so also are A and* B^c.

Proof.

$$\begin{aligned}\mathbb{P}(A \cap B^c) =& \mathbb{P}(A) - \mathbb{P}(A \cap B) \quad (\text{since } \mathbb{P}(A \cap B^c) + \mathbb{P}(A \cap B) = \mathbb{P}(A)) \\ =& \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) \\ =& \mathbb{P}(A)(1 - \mathbb{P}(B)) = \mathbb{P}(A)\mathbb{P}(B^c)\end{aligned}$$

□

The definition of the independence of random variables relies on the following definition:

Definition 2.3. *A σ-field $\mathcal{F}$ is independent of a σ-field $\mathcal{G}$ if each $A \in \mathcal{F}$ is independent of each $B \in \mathcal{G}$.*

Two random variables X and Y are said to be independent if and only if the value of X has no influence on the value of Y and vice versa. In terms of σ-fields it looks like this:

Definition 2.4. *Random variables X and Y are independent if the σ-fields generated by them are independent.*

Here is an example of independent continuous random variables. Imagine that we are throwing darts into a square board with dimensions 1x1 and assume that every time we necessarily hit the square. The outcomes are $\{(x,y)|x \in [0,1], y \in [0,1]\}$. Since the probability of separate outcomes is zero, we assign probabilities for subsets of a unit interval. Let X be the random variable that represents the x-coordinate of the dart after the throw, while Y represents the y-coordinate. Then probability that X "hits" the interval $[a,b] \subset [0,1]$ (denote this event as A) as well as the probability that Y "hits" the interval $[c,d]$ (denote this event as B) are areas of the corresponding "stripes", namely $(b-a)\cdot 1$ and $(d-c)\cdot 1$ (Figure (2.1)). The probability that darts hits the board within the limits of both intervals $[a,b]$ and $[c,d]$ is the area of the blue rectangle: $(b-a)\cdot(d-c)$. In other words is $\mathbb{P}(A \cap B) = (b-a)\cdot(d-c) = (b-a)\cdot 1\cdot(d-c)\cdot 1 = \mathbb{P}(A)\cdot\mathbb{P}(B)$, so X and Y are independent.

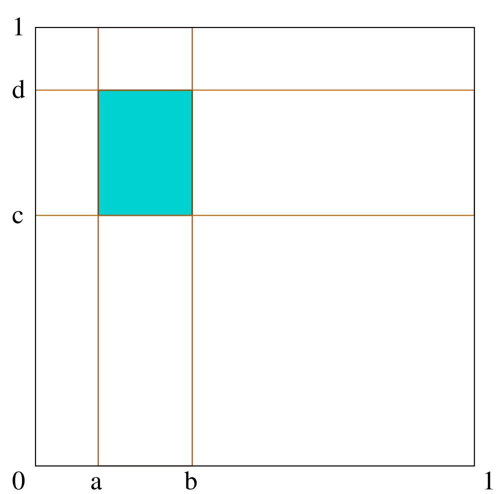

Figure 2.1: *Square dartboard*

Theorem 2.5. *If X, Y are independent random variables, f,g are Borel measurable, then f(X) and g(Y) are independent.*

Proof. Let X, Y, f, g be as above. Choose $A \subset \mathbb{R}$, $B \subset \mathbb{R}$ Borel measurable. Then

$$\begin{aligned}\mathbb{P}((f\circ X)^{-1}(A) \cup (g\circ Y)^{-1}(B)) &= \mathbb{P}(X^{-1}(f^{-1}(A)) \cup Y^{-1}(g^{-1}(B))) = \quad \text{(see Lemma 1.12)}\\ &= \mathbb{P}(X^{-1}(f^{-1}(A))) \cdot \mathbb{P}(Y^{-1}(g^{-1}(B))) = \\ &= \mathbb{P}((f\circ X)^{-1}(A)) \cdot \mathbb{P}((g\circ Y)^{-1}(B))\end{aligned}$$

□

Theorem 2.6. *If X, Y are independent random variables, then $\mathbb{E}(XY) = (\mathbb{E}X)(\mathbb{E}Y)$.*

Proof. Let X, Y be as above. We denote $(\Omega, \mathcal{F}, \mathbb{P})$ as a probability space. Without loss of generality assume $X, Y \geq 0$. Define $S_k := \{m2^{-k}\}_{m=1}^{2^k k}$. Consider the following sequences of random variables

$$X_k(\omega) := \sum_{m=1}^{2^k(k-1)} m2^{-k}\chi_{X^{-1}([m2^k,(m+1)2^k))}(\omega) + 2^k k\chi_{X^{-1}[2^k k,\infty)}(\omega)$$
$$Y_k(\omega) := \sum_{n=1}^{2^k(k-1)} n2^{-k}\chi_{Y^{-1}([n2^k,(n+1)2^k))}(\omega) + 2^k k\chi_{Y^{-1}[2^k k,\infty)}(\omega)$$

so that $X_k \nearrow X$, $Y_k \nearrow Y$, and consequently $X_k Y_k \nearrow XY$. Say:

$$X_k(\omega) := \sum_{m=1}^{N} a_{m,k}\chi_{X^{-1}(S_{m,k})}(\omega)$$
$$Y_k(\omega) := \sum_{n=1}^{N} b_{n,k}\chi_{Y^{-1}(S_{n,k})}(\omega)$$

Now

$$\begin{aligned}\int X(\omega)\cdot Y(\omega)\, d\mathbb{P}(\omega) &= \int \lim_{k\to\infty} X_k(\omega)\cdot Y_k(\omega)\, d\mathbb{P}(\omega) = \lim_{k\to\infty}\int X_k(\omega)\cdot Y_k(\omega)\, d\mathbb{P}(\omega) = \quad \text{(by MCT)}\\ &= \lim_{k\to\infty}\int \sum_{m=1}^{N} a_{m,k}\chi_{X^{-1}(S_{m,k})} \cdot \sum_{n=1}^{N} b_{n,k}\chi_{X^{-1}(S_{n,k})}\, d\mu = \\ &= \lim_{k\to\infty}\sum_{m=1}^{N}\sum_{n=1}^{N} a_{m,k}b_{n,k}\int \chi_{X^{-1}(S_{m,k})}\cdot\chi_{Y^{-1}(S_{n,k})} = I\end{aligned}$$

Note that $\chi_A \cdot \chi_B \Leftrightarrow (\omega \in A)\&(\omega \in B) \Leftrightarrow \chi_{A\cap B}$, so

$$
\begin{aligned}
I &= \lim_{k\to\infty} \sum_{m=1}^{N}\sum_{n=1}^{N} a_{m,k}b_{n,k} \int \chi_{(X^{-1}(S_{m,k})\cap Y^{-1}(S_{n,k}))} = \\
&= \lim_{k\to\infty} \sum_{m=1}^{N}\sum_{n=1}^{N} a_{m,k}b_{n,k}\, \mathbb{P}(X^{-1}(S_{m,k}) \cap Y^{-1}(S_{n,k})) = \\
&= \lim_{k\to\infty} \sum_{m=1}^{N}\sum_{n=1}^{N} a_{m,k}b_{n,k}\, \mathbb{P}(X^{-1}(S_{m,k})) \cdot \mathbb{P}(Y^{-1}(S_{n,k})) = \\
&= \lim_{k\to\infty} \sum_{m=1}^{N}\sum_{n=1}^{N} a_{m,k}b_{n,k} \int \chi_{X^{-1}(S_{m,k})} \cdot \int \chi_{Y^{-1}(S_{n,k})} = \\
&= \lim_{k\to\infty} \int \sum_{m=1}^{N} a_{m,k}\chi_{X^{-1}(S_{m,k})} \cdot \int \sum_{n=1}^{N} b_{n,k}\chi_{Y^{-1}(S_{n,k})} = \\
&= \lim_{k\to\infty} \int X_k(\omega)\, d\mathbb{P}(\omega) \cdot \int Y_k(\omega)\, d\mathbb{P}(\omega) = \\
&= \lim_{k\to\infty} \int X_k(\omega)\, d\mathbb{P}(\omega) \cdot \lim_{k\to\infty} \int Y_k(\omega)\, d\mathbb{P}(\omega) = \\
&= \int X(\omega)\, d\mathbb{P}(\omega) \cdot \int Y(\omega)\, d\mathbb{P}(\omega)
\end{aligned}
$$

□

Definition 2.7. *The characteristic function of a random variable X is the Fourier transform of its law: $\int e^{iux} d\,\mathbb{P}_X = \mathbb{E}(e^{iuX})$.*

The Fourier transform is a fundamental notion in probability and analysis. We will see that the Fourier transform of a measure determines that measure. We will define Fourier transform as well as its basic properties in the next chapter.

Definition 2.8. *The joint characteristic function of random variables X and Y is the following expectation: $\mathbb{E}(e^{i(uX+vY)})$.*

Note that if X and Y are independent random variables, then e^{iuX} and e^{ivY} are also independent by Theorem 2.5. Therefore $\mathbb{E}(e^{i(uX+vY)}) = \mathbb{E}(e^{iuX})\mathbb{E}(e^{ivY})$ by Theorem 2.6. In fact the converse also holds. This is a major result and the proof will take up the remainder of this chapter.

Definition 2.9. *If $T : (X, \mathcal{A}) \to (Y, \mathcal{D})$ is a measurable function and μ is a measure on $(X, \mathcal{A})$, then $T_*\mu : \mathcal{D} \to [0, \infty)$, $(T_*\mu)(B) = \mu(T^{-1}(B))$ is the push forward measure.*

Lemma 2.10. $\chi_{T^{-1}(B)} = \chi_B \circ T$.

Proof. Let $x \in X$. Then either $x \in T^{-1}(B)$ or $x \notin T^{-1}(B)$.

- case 1: $x \in T^{-1}(B) \Rightarrow T(x) \in B$. So $(\chi_B \circ T)(x) = \chi_B(T(x)) = 1 = \chi_{T^{-1}(B)}(x)$.
- case 2: $x \notin T^{-1}(B) \Rightarrow T(x) \notin B$. So $(\chi_B \circ T)(x) = \chi_B(T(x)) = 0 = \chi_{T^{-1}(B)}(x)$.

$\square$

Lemma 2.11. *For any $f \in L^1(T_*\mu)$*

$$\int_Y f dT_*\mu = \int_X f \circ T d\mu$$

Proof. case 1 (f is simple): $f = \sum a_k \chi_{Ak}$.

$$\begin{aligned}
\int_Y \sum a_k \chi_{Ak} d(T_*\mu) &= \sum a_k T_*\mu(A_k) \\
&= \sum a_k \mu(T^{-1}(A_k)) \quad \text{(by definition of push forward measure)} \\
&= \int_X \sum a_k \chi_{T^{-1}(A_k)} d\,\mu \\
&= \int_X \sum a_k \chi_{A_k} \circ T d\mu \quad \text{(by Lemma (2.10))}
\end{aligned}$$

case 2 ($f \geq 0$): there exists sequence of simple functions $s_n \nearrow f$

$$\begin{aligned}
\int_Y f dT_*\mu &= \int_Y \lim_{n\to\infty} s_n dT_*\mu = \lim_{n\to\infty} \int_Y s_n dT_*\mu \quad \text{(by MCT)} \\
&= \lim_{n\to\infty} \int_X s_n \circ T d\mu = \int_X \lim_{n\to\infty} s_n \circ T d\mu \quad \text{(by MCT)} \\
&= \int_X f \circ T d\mu
\end{aligned}$$

case 3 ($f \in \mathbb{R}$-valued): first define $f = f_+ - f_-$ where

$$f_+(t) = \begin{cases} f(t) & \text{if } f(t) > 0 \\ 0 & \text{otherwise} \end{cases} \qquad f_-(t) = \begin{cases} f(t) & \text{if } f(t) < 0 \\ 0 & \text{otherwise} \end{cases}$$

With this notation in mind:

$$\begin{aligned}
\int f d\, T_*\mu &= \int_Y f_+ d\, T_*\mu - \int_Y f_- d\, T_*\mu \\
&= \int_X f_+ \circ T\, d\mu - \int_X f \circ T\, d\mu \quad \text{(by previous case)} \\
&= \int_X (f \circ T)_+\, d\mu - \int_X (f \circ T)_-\, d\mu = \int_X f \circ T\, d\mu
\end{aligned}$$

case 4 ($f \in \mathbb{C}$-valued): $f = p + iq$, then

$$\int f d\, T_*\mu = \int (p + iq) d\, T_*\mu = \int p d\, T_*\mu + i \int q d\, T_*\mu$$

$\square$

Definition 2.12. *Given random variables X, Y, $\mathbb{P}_{(X,Y)} = (X,Y)_*\mathbb{P}$ where $\mathbb{P}_{(X,Y)}(A \times B) = \mathbb{P}(X^{-1}(A) \cap Y^{-1}(B))$.*

Theorem 2.13. *If random variables X, Y satisfy $\mathbb{E}(e^{iuX+ivY}) = \mathbb{E}(e^{iuX})\mathbb{E}(e^{ivY})$ for all u and v, then X and Y are independent.*

Proof. Denote $\mathbb{P}_X$, $\mathbb{P}_Y$ probability measures of X and Y correspondingly. Then by Theorem (1.25) there exist random variables X'' and Y'' such that $X'' : [0,1] \to \mathbb{R}$, $Y'' : [0,1] \to \mathbb{R}$ with $\mathbb{P}_{X''} = \mathbb{P}_X$, $\mathbb{P}_{Y''} = \mathbb{P}_Y$. Now consider probability space ($[0,1]^2$, Borel σ-algebra, Lebesgue measure with total mass 1) and set $X'(s,t) = X''(s)$, $Y'(s,t) = Y''(t)$, so $\mathbb{P}_{X'} = \mathbb{P}_{X''} = \mathbb{P}_X$ and $\mathbb{P}_{Y'} = \mathbb{P}_{Y''} = \mathbb{P}_Y$. So X', Y' are our prototypes of independent random variables (X', Y') are independent by the example on page 13) and recall $\mathbb{P}_{(X,Y)} = (X,Y)_*\mathbb{P}$, where $(X,Y) : \Omega \to \mathbb{R}^2$, $\omega \mapsto (X(\omega), Y(\omega))$. Also note that

$$\mathbb{E}(X) = \int_\Omega X(\omega)\, d\mathbb{P}(\omega) = \int_\mathbb{R} x\, d\mathbb{P}_X(x) = \int_\mathbb{R} x\, d\mathbb{P}_{X'}(x) = \int_{[0,1]} X'(t)\, d\mu(t) = \mathbb{E}(X').$$

Now consider

$$\begin{aligned}
(\mathcal{F}\,\mathbb{P}_{X,Y})(u,v) &= \int_{\mathbb{R}^2} e^{i(ux+vy)}\, d\mathbb{P}_{(X,Y)}(x,y) = \\
&= \int_{\mathbb{R}^2} e^{i(uX+vY)}\, d(X,Y)_*\mathbb{P} = \quad \text{(by Lemma (2.11))} \\
&= \int_\Omega e^{i(uX(\omega)+vY(\omega))}\, d\mathbb{P}(\omega) = \\
&= \mathbb{E}(e^{i(uX+vY)}) = \mathbb{E}(e^{iX})\mathbb{E}(e^{iY}) = \\
&= \mathbb{E}(e^{iX'})\mathbb{E}(e^{iY'}) = \mathbb{E}(e^{i(uX'+vY')}) = \quad \text{(since } X', Y' \text{are independent)} \\
&= \mathbb{E}(e^{i(uX'+vY')}) = \int_{[0,1]^2} e^{i(uX'(s,t)+vY'(s,t))}\, d\mu(s,t) = \\
&= \int_{\mathbb{R}^2} e^{i(ux+vy)}\, d\mathbb{P}_{(X',Y')}(x,y) = \quad \text{(by Lemma 2.11)} \\
&= (\mathcal{F}\,\mathbb{P}_{X',Y'})(u,v)
\end{aligned}$$

therefore $\mathbb{P}_{(X,Y)} = \mathbb{P}_{(X',Y')}$ by Lemma (3.15). We will prove Lemma (3.15) in the next chapter. Now use the definition of $\mathbb{P}_{(X,Y)}$:

$$\begin{aligned}
\mathbb{P}(X^{-1}(A) \cap Y^{-1}(B)) &= \mathbb{P}_{(X,Y)}(A \times B) = \quad \text{(definition of } \mathbb{P}_{(X,Y)}) \\
&= \mathbb{P}_{(X',Y')}(A \times B) = \quad \text{(by the result above)} \\
&= \mathbb{P}((X')^{-1}(A) \cap (Y')^{-1}(B)) = \quad \text{(definition of } \mathbb{P}_{(X',Y')})) \\
&= \mathbb{P}((X')^{-1}(A)) \cdot \mathbb{P}((Y')^{-1}(B)) = \quad (X', Y' \text{ are independent}) \\
&= \mathbb{P}_{X'}(A) \cdot \mathbb{P}_{Y'}(B) = \\
&= \mathbb{P}_X(A) \cdot \mathbb{P}_Y(B) = \\
&= \mathbb{P}(X^{-1}(A)) \cdot \mathbb{P}(Y^{-1}(B))
\end{aligned}$$

so X, Y are independent by the definition of independence of random variables. $\square$

Chapter 3

Fourier Transform

The Fourier transform will be used to prove the law of large numbers and the central limit theorem in the next chapter. In this chapter we define the Fourier transform for functions and for measures, prove properties for functions and then extend those properties to measures.

Definition 3.1. *If $f \in L^1(\mathbb{R})$, then the Fourier transform is $\mathcal{F}(f)(\omega) = \widehat{f}(\omega) = \int_{\mathbb{R}} e^{i\omega x} f(x)\, dx$.*

Remark: Recall that the Fourier transform of a measure μ is $(\mathcal{F}\mu)(\omega) = \int_{\mathbb{R}} e^{i\omega x}\, d\mu(x)$. For positive $f \in L^1(\mathbb{R})$ if one sets

$$\mu_f(A) = \int_A f(x)\, dx \quad \text{ then } \quad \widehat{f}(\omega) = (\mathcal{F}\mu_f)(\omega)$$

Example 3.1
If $f(x) = \frac{1}{2\pi} e^{-\frac{1}{2}\frac{x^2}{\sigma^2}}$, then $\mathcal{F}(f)(u) = \frac{1}{\sqrt{2\pi}} \sigma e^{-\frac{1}{2}\sigma^2 u^2}$.

$$\begin{aligned}
\mathcal{F}(\frac{1}{2\pi} e^{-\frac{1}{2}\frac{x^2}{\sigma^2}})\,(u) =& \frac{1}{2\pi} \int_{\mathbb{R}} e^{iux - \frac{1}{2}\frac{x^2}{\sigma^2}}\, dx \\
=& \frac{1}{2\pi} \int_{\mathbb{R}} e^{-\frac{1}{2\sigma^2}(x^2 - 2\sigma^2 iux + \sigma^4 u^2 - \sigma^4 u^2)}\, dx \\
=& \frac{1}{2\pi} e^{-\frac{1}{2}\sigma^2 u^2} \int_{\mathbb{R}} e^{-\frac{1}{2\sigma^2}(x - \sigma^2 iu)^2}\, dx \\
=& \frac{1}{2\pi} \sigma \sqrt{2\pi} e^{-\frac{1}{2}\sigma^2 u^2} = \frac{1}{\sqrt{2\pi}} \sigma e^{-\frac{1}{2}\sigma^2 u^2}
\end{aligned}$$

One may set $k = \frac{1}{\sigma}$ to obtain $\frac{1}{2\pi} e^{-\frac{1}{2}\frac{x^2}{\sigma^2}} = \frac{1}{2\pi} e^{-\frac{1}{2}k^2 x^2} = f(x)$ and $\mathcal{F}(f)(u) = \frac{1}{k\sqrt{2\pi}} e^{-\frac{1}{2}\frac{u^2}{k^2}}$.

Our first goal is to establish the Fourier inversion formula for functions. The following lemma is a key element in the proof of the Fourier inversion formula for functions. It is sometimes called the Fourier multiplication theorem.

Lemma 3.2. *If $f, g \in L^1$, then $\int_{\mathbb{R}} \widehat{f}(y) g(y) dy = \int_{\mathbb{R}} \widehat{g}(x) f(x) dx$*

Proof.

$$\begin{aligned}
\int_{\mathbb{R}} \widehat{f}(y)g(y)dy = \\
&= \int_{\mathbb{R}} (\int_{\mathbb{R}} e^{iyx} f(x)dx)g(y)dy \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} e^{iyx} f(x)g(y)dydx \\
&\quad \text{(Note: by Tonelli; } |e^{iyx} f(x)g(y)| \le |f(x)||g(y)| \in L^1 \text{ as } f, g \in L^1) \\
&= \int_{\mathbb{R}} (\int_{\mathbb{R}} e^{iyx} g(y)dy) f(x)dx \quad \text{(by Fubini)} \\
&= \int_{\mathbb{R}} \widehat{g}(x) f(x)dx
\end{aligned}$$

□

We now prove the inversion formula for the function at zero.

Lemma 3.3. *If $g \in C^0$, $\widehat{g}$ is defined and $\widehat{g} \in L^1$, then $g(0) = \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{g}(u)du$.*

Proof.

$$\begin{aligned}
g(0) &= \lim_{k \to 0} g(kz) \\
&= \lim_{k \to 0} g(kz) \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\, dz \\
&= \lim_{k \to 0} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} g(kz)\, dz \\
&= \lim_{k \to 0} \int_{\mathbb{R}} \frac{1}{k\sqrt{2\pi}} e^{-\frac{x^2}{2k^2}} g(x)\, dx \quad (x = kz, z = \frac{x}{k}, dz = \frac{dx}{k}) \\
&= \lim_{k \to 0} \frac{1}{2\pi} \int_{\mathbb{R}} e^{-\frac{k^2u^2}{2}} \widehat{g}(u)\, du \quad (g \in L^1, \text{ Lemma (3.2) and example on page 18}) \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} \lim_{k \to 0} e^{-\frac{k^2u^2}{2}} \widehat{g}(u)\, du \quad (\text{DCT: } |e^{-\frac{k^2u^2}{2}} \widehat{g}(u)| \le |\widehat{g}(u)|, \widehat{g} \in L^1) \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{g}(u)\, du
\end{aligned}$$

□

We can now prove the inversion formula for functions.

Theorem 3.4. *If $h \in C^0, L^1$, $\widehat{h}$ is defined and is in L^1, then $h(y) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iuy} \widehat{h}(u)du$.*

Proof. Set $g(x) = h(x+y)$. Clearly $g \in C^0$ as h is. Now

$$\begin{aligned}
\widehat{g}(u) &= \int_{\mathbb{R}} e^{iux} g(x)\, dx \\
&= \int_{\mathbb{R}} e^{iux} h(x+y)\, dx \\
&= \int_{\mathbb{R}} e^{iu(z-y)} h(z)\, dz \quad (z = x+y, x = z-y, dx = dz) \\
&= e^{-iuy} \int_{\mathbb{R}} e^{iuz} h(z)\, dz \\
&= e^{-iuy} \widehat{h}(u)
\end{aligned}$$

That is $|\widehat{g}(u)| = |e^{-iuy}\widehat{h}(u)| = |\widehat{h}(u)|$, so $\widehat{g} \in L^1$ as $\widehat{h} \in L^1$. Thus

$$\begin{aligned}
h(y) = g(0) &= \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{g}(u)\, du \quad (\text{by Lemma (3.3) as } g \in C^1, \widehat{g} \in L^1) \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iuy} \widehat{h}(u)\, du
\end{aligned}$$

□

There is a multiplication that is closely associated with the Fourier transform. It is called convolution.

Definition 3.5. *If two real-valued function f, g are sufficiently summable, then convolution is given by*

$$(f * g)(x) = \int_{\mathbb{R}} f(x-y)g(y)dy$$

Convolution is also defined between a function and a measure.

Definition 3.6. *The convolution of a function and a measure is given by*

$$(f * \mu)(x) = \int_{\mathbb{R}} f(x-y)d\mu(y)$$

We use convolution to approximate measures by smooth measures and to translate results about function to results about measures.

Definition 3.7. *We say the sequence of measures μ_n converges to measure μ narrowly (weakly) if and only if $\int f(x)d\mu_n(x) \to \int f(x)d\mu(x)$ for every $f \in L^\infty \cap C^0$.*

Let ρ_1 be a smooth positive bump function with mass 1 ($\int \rho_1(x)dx = 1$) such that $\rho_1 = 0$ for $|x| > 1$. Set $\rho_n(x) := n\rho_1(nx)$, note that $\rho_n(x) = 0$ for $|x| > \frac{1}{n}$, and ρ_n still has mass 1 (Figure (3.1)).

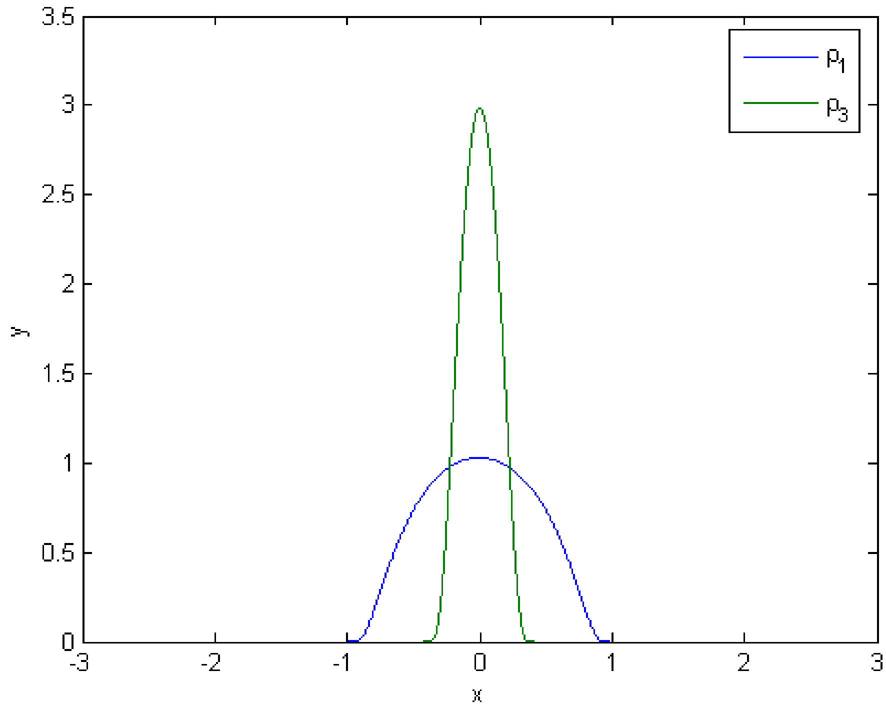

Figure 3.1: *Smooth positive bump functions*

Lemma 3.8. *If μ is a finite measure, then $\mu_n = \rho_n * \mu dx \to \mu$ narrowly.*

Proof. We need to show that $\lim_{n\to\infty} \int f(x) * \mu_n(x) dx = \int f(y) d\mu(y)$ for bounded continuous functions. We first prove it for non-negative bounded continuous functions, say $0 \leq f(x) \leq M$. Notice that $\int f(x)\rho_n(x-y)dx \leq \int M\rho_n(x-y)dx \leq M \int \rho_n(x-y)dx \leq M$, since $\rho_n(x)$ has mass 1. So $M \in L^1(\mu)$ as μ is a finite measure. For such functions we have

$$\begin{aligned}
\lim_{n\to\infty} \int f(x) d\mu_n(x) dx &= \lim_{n\to\infty} \int f(x) \int \rho_n(x-y) d\mu(y) dx \quad &\text{(def of } \mu_n) \\
&= \lim_{n\to\infty} \int \int f(x)\rho_n(x-y) dx \, d\mu(y) \quad &\text{(by Tonelli's theorem)} \\
&= \int \lim_{n\to\infty} \int f(x)\rho_n(x-y) dx \, d\mu(y) \quad &\text{(by DCT as } M \in L^1(\mu)) \\
&= \int f(y) \, d\mu(y)
\end{aligned}$$

We write a general bounded continuous function $f(x) = f_+(x) - f_-(x)$, where

$$f_+(x) = \begin{cases} f(x) & \text{if } f(x) \geq 0 \\ 0 & \text{else} \end{cases} \qquad f_-(x) = \begin{cases} -f(x) & \text{if } f(x) \leq 0 \\ 0 & \text{else} \end{cases}$$

and use linearity. $\square$

Lemma 3.9. *If $\mu, (\mu_n)_{k\geq 1}$ are probability measures with cumulative distribution functions $F, (F_n)_{k\geq 1}$ such that $F_n \to F$ on a dense subset of $\mathbb{R}$ say $\triangle$ then $\mu_n \to \mu$ narrowly.*

Proof. There exist $r, s \in \triangle$ such that $1 - \mu((r,s]) = 1 - F(s) + F(r) \leq \epsilon$. Also since $F_n \to F$ on $\triangle$ by hypothesis, there exists an N_1 such that for $n \geq N_1$ $1 - \mu_n((r,s]) = 1 - F_n(s) + F_n(r) \leq \epsilon$. Since $[r,s]$ is a closed (compact) interval and f is continuous, we know f is uniformly continuous on $[r,s]$; thus there exists a finite number of points $r = r_0 < r_1, \ldots, r_k = s$ such that

$$|f(x) - f(r_j)| \leq \epsilon \quad \text{if } r_{j-1} \leq x \leq r_j \tag{3.1}$$

and each of the r_j are in $\triangle$, $1 \leq j \leq k$. Next we set

$$g(x) = \sum_{j=1}^{k} f(r_j)\chi_{(r_{j-1},r_j]}(x) \tag{3.2}$$

and by (3.1) we have $|f(x) - g(x)| \leq \epsilon$ on $(r,s]$. If $\alpha = \sup_x |f(x)|$, we obtain

$$\begin{aligned}
\left|\int_{\mathbb{R}} f(x)d\mu_n(x) - \int_{\mathbb{R}} g(x)d\mu_n(x)\right| &= \left|\int_{\mathbb{R}} (f(x) - g(x))d\mu_n(x)\right| \\
&= \left|\int_{(r,s]^c} (f(x) - g(x))d\mu_n(x) + \int_{(r,s]} (f(x) - g(x))d\mu_n(x)\right| \\
&\leq \left|\int_{(r,s]^c} (f(x) - g(x))d\mu_n(x)\right| + \left|\int_{(r,s]} (f(x) - g(x))d\mu_n(x)\right| \\
&\leq \int_{(r,s]^c} |f(x) - g(x)|d\mu_n(x) + \int_{(r,s]} |f(x) - g(x)|d\mu_n(x) \\
&\leq \int_{(r,s]^c} |f(x)| + |g(x)|d\mu_n(x) + \int_{(r,s]} \epsilon d\mu_n(x) \\
&\leq 2\alpha\mu_n((r,s]^c) + \epsilon\mu_n((r,s]) \\
&\leq 2\alpha 2\epsilon + \epsilon 1 = (4\alpha + 1)\epsilon
\end{aligned} \tag{3.3}$$

Similarly

$$\left|\int_{\mathbb{R}} f(x)d\mu(x) - \int_{\mathbb{R}} g(x)d\mu(x)\right| \leq (4\alpha + 1)\epsilon \tag{3.4}$$

Using definition of g (3.2)

$$\begin{aligned}\int_{\mathbb{R}} g(x)d\mu(x) &= \int_{\mathbb{R}} \sum_{j=1}^{k} f(r_j)\chi_{(r_{j-1},r_j]}(x)d\mu(x)\\ &= \sum_{j=1}^{k} f(r_j)\mu((r_{j-1},r_j])(x)\\ &= \sum_{j=1}^{k} f(r_j)(F(r_{j-1})-F(r_j))\end{aligned}$$

and analogously

$$\int_{\mathbb{R}} g(x)d\mu_n(x) = \sum_{j=1}^{k} f(r_j)(F_n(r_{j-1})-F_n(r_j))$$

Since all the r_j's are in $\triangle$, we have $\lim_{n\to\infty} F_n(r_j) = F(r_j)$ for each j. And since there are only finite number of r_j's, we know there exists an N_2 such that for $n \geq N_2$, $|\int_{\mathbb{R}} g(x)d\mu_n(x) - \int_{\mathbb{R}} g(x)d\mu_n(x)| < \epsilon$. Using this result as well as (3.3) and (3.4) we get

$$\begin{aligned}&\left|\int f(x)d\mu_n(x) - \int f(x)d\mu(x)\right| =\\ =&\left|\int f(x)d\mu_n(x) - \int g(x)d\mu_n(x) + \int g(x)d\mu_n(x) - \int g(x)d\mu(x) + \int g(x)d\mu(x) - \int f(x)d\mu(x)\right|\\ =&\left|\int f(x)d\mu_n(x) - \int g(x)d\mu_n(x)\right| + \left|\int g(x)d\mu_n(x) - \int g(x)d\mu(x)\right| + \left|\int g(x)d\mu(x) - \int f(x)d\mu(x)\right|\\ \leq&(4\alpha+1)\epsilon + \epsilon + (4\alpha+1)\epsilon = 8\alpha\epsilon + 3\epsilon\end{aligned}$$

Since ϵ as arbitrary, $\int f(x)d\mu_n(x) \to \int f(x)d\mu(x)$ for all bounded, continuous f. □

Lemma 3.10. *Let $(\mu_n)_{n\geq 1}$ be a sequence of probability measures on $\mathbb{R}$ and suppose*

$$\lim_{m\to\infty} \sup_n \mu_n([-m,m]^c) = 0$$

Then there exists a probability measure μ such that $(\mu_n)_{n\geq 1}$ converges narrowly to μ.

Proof. Let $F_n(x) = \mu_n((-\infty,x])$. Note that $(F_n(x))_{n\geq 1}$ is a bounded sequence of real numbers since for each $x \in \mathbb{R}, 0 \leq F_n(x) \leq 1$ for all n. Then by the Bolzano-Weierstrass theorem there always exists a subsequence n_k such that $(F_n(x))_{n\geq 1}$ converges where n_k depends on x.

Let $r_1, r_2, \ldots, r_j, \ldots$ be a sequence of rational numbers. For r_1, there exists a subsequence $n_{1,k}$ of n such that the limit exists and we set $X(r_1) = \lim_{k\to\infty} F_{n_{1,k}}(r_1)$. For r_2, there exists a sub-subsequence $n_{2,k}$ of $n_{1,k}$ such that the limit exists and we set $X(r_2) = \lim_{k\to\infty} F_{n_{2,k}}(r_2)$.

We continue this way: for r_j, let $n_{j,k}$ be a sequence of $n_{j-1,k}$ such that the limit exists and we set $X(r_j) = \lim_{k\to\infty} F_{n_{j,k}}(r_j)$. We then form just one subsequence by taking $n_k := n_{k,k}$. Thus for f_j, we have $X(r_j) = \lim_{k\to\infty} F_{n_k}(r_j)$ since n_k is a subsequence of $n_{j,k}$ once $k \geq j$.

Next, we set $F(x) = \inf_{y>x} X(y)$. Function F is right-continuous by construction as well as non-decreasing (since X defined on Q is non-decreasing).

Let $\epsilon > 0$. By hypothesis there exists an m such that $\mu_n([-m,m]^c) \leq \epsilon$ for all n simultaneously. Therefore $F_n(x) \leq \epsilon$ if $x < -m$ and $F_n(x) \geq 1-\epsilon$ if $x > m$, therefore we have same for X and finally

$$\left.\begin{array}{rl} F(x) \leq \epsilon & \text{if } x < -m \\ F(x) \geq 1-\epsilon & \text{if } x \geq m \end{array}\right\}$$

Since $0 \leq F(x) \leq 1$, F is right continuous and non-decreasing, F is a distribution function corresponding to a probability measure μ.

Finally, suppose x is such that $F(x-) = \lim_{y\to x^-} F(y) = F(x)$. For $\epsilon > 0$ there exists $y, z \in \mathbb{Q}$ with $y < x < z$ and $F(x) - \epsilon \leq X(y) \leq F(x) \leq X(x) \leq F(x) + \epsilon$. Thus for large enough k,

$$F(x) - 2\epsilon \leq F_{n_k}(y) \leq F_{n_k}(x) \leq F_{n_k}(z) \leq F(x) + 2\epsilon$$

By the inequalities above

$$F(x) - 2\epsilon \leq F(y) \leq \liminf_{k\to\infty} F_{n_k}(x) \leq \limsup_{k\to\infty} F_{n_k}(x) \leq F(z) \leq F(x) + 2\epsilon$$

and by the squeeze theorem the lim inf and lim sup above must be equal and equal to $\lim_{k\to\infty} F_{n_k}(x) = F(x)$. Thus μ_{n_k} converges narrowly to μ by Lemma (3.9). $\square$

The relationship between Fourier transforms and convolution is given in the the next lemma.

Lemma 3.11. *For sufficiently smooth functions f, g we have $\mathcal{F}((f*g)(x))(u) = \mathcal{F}(f)(u) \cdot \mathcal{F}(g)(u)$*

Proof.

$$\begin{aligned} \mathcal{F}((f*g)(x))(u) =& \mathcal{F}(\int_{\mathbb{R}} f(x-y)g(y)dy)(u) = \\ =& \int_{\mathbb{R}} e^{iux}(\int_{\mathbb{R}} f(x-y)g(y)dy)dx \\ =& \int_{\mathbb{R}}\int_{\mathbb{R}} e^{iux} f(x-y)g(y)dydx =: J \end{aligned}$$

Denote $z = x - y$, so $x = z + y$ and so

$$\begin{aligned} J =& \int_{\mathbb{R}}\int_{\mathbb{R}} e^{iuz} f(z)dz e^{iuy} g(y)dy \\ =& \mathcal{F}(f)(u) \cdot \mathcal{F}(g)(u) \end{aligned}$$

$\square$

We will use this for the convolution of a bump function and a finite measure in the proof of the main theorem of this chapter.

Lemma 3.12. $\mathcal{F}(\rho * \mu)(u) = \widehat{\rho}(u)\mathcal{F}(\mu)(u)$

Proof.

$$\begin{aligned}
\mathcal{F}(\mu)(u) = \mathcal{F}(\rho * \mu)(u) &= \int e^{iux} \int \rho(x-y)d\mu(y)dx \\
&= \int\int e^{iux}\rho(x-y)dxd\mu(y) \quad \text{(by Tonelli's theorem)} \\
&= \int e^{iuy}\int e^{iuz}\rho(z)dzd\mu(y) \quad (z = x-y; x = z+y) \\
&= \widehat{\rho}(u)\mathcal{F}\mu
\end{aligned}$$

□

We need to know how smooth and summable $\rho_n * \mu$ is for our main theorem.

Lemma 3.13. *If μ is a finite measure and ρ_n is a family of bump functions, then $\rho_n * \mu$ is bounded and smooth i.e. infinitely differentiable. Furthermore all derivatives tend to zero at infinity.*

Proof. Indeed,

$$|\rho_n * \mu(x)| \leq \int \rho_n(x-y)d\mu(y) \leq \int \sup_{x\in\mathbb{R}} |\rho_n(x-y)|d\mu(y) \leq \sup_{z\in\mathbb{R}} |\rho_n(z)|$$

since $\rho_n(x-y)$ is continuous with compact support.

$$\begin{aligned}
|\rho_n * \mu(x) - \rho_n * \mu(z)| &= \left| \int \left(\rho_n(x-y) - \rho_n(z-y)\right) d\mu(y) \right| \\
&\leq \int \left|\rho_n(x-y) - \rho_n(z-y)\right| d\mu(y)
\end{aligned}$$

Since ρ_n is infinitely differentiable with compact support, it is uniformly continuous. So for any $\epsilon > 0$ there exists $\delta > 0$ such that $|x-z| < \delta$ and $\rho_n * \mu$ is continuous implies $|\rho_n(x-y) - \rho_n(z-y)| < \frac{\epsilon}{\int d\mu(y)}$. Thus $|\rho_n * \mu(x) - \rho_n * \mu(z)| < \epsilon$.

Note that $\left|\frac{\rho_n(x+h-y) - \rho_n(x-y)}{h}\right|$ is dominated by a function in L^1 because

$$\begin{aligned}
\left|\frac{\rho_n(x+h-y) - \rho_n(x-y)}{h}\right| &= |\rho_n'(c)| \quad \text{(by MVT as } \rho_n \in C_0^\infty) \\
&\leq \sup_{c\in\mathbb{R}} |\rho_n'(c)| \in L^1 \quad \text{(since } \mu \text{ is finite)}
\end{aligned}$$

So using the above observation

$$\begin{aligned}
\frac{d}{dx}(\rho_n * \mu(x)) &= \lim_{h\to 0}\frac{1}{h}(\rho_n * \mu(x+h) - \rho_n * \mu(x)) \\
&= \lim_{h\to 0}\int \frac{1}{h}(\rho_n(x+h-y) - \rho_n(x-y))d\mu(y) \\
&= \int \lim_{h\to 0}\frac{1}{h}(\rho_n(x+h-y) - \rho_n(x-y))d\mu(y) \quad \text{(by DCT)} \\
&= \int \rho_n'(x)\, d\mu(y) = \rho_n' * \mu
\end{aligned}$$

Similarly $\frac{d^k}{dx^k}(\rho_n * \mu(x)) = \rho_n^{(k)} * \mu(x)$. □

This is what we mean when we say $\rho_n * \mu$ is smooth.

Lemma 3.14. $\lim_{x\to\infty} \rho_n * \mu(b) = 0$

Proof. Note that

$$\mu(\mathbb{R}) = \mu\left(\bigcup_{k=-\infty}^{+\infty}[2k, 2k+2)\right) = \sum_{k=-\infty}^{+\infty}\mu([2k,2k+2))$$

$$\mu(\mathbb{R}) = \mu\left(\bigcup_{k=-\infty}^{+\infty}[2k-1, 2k+1)\right) = \sum_{k=-\infty}^{+\infty}\mu([2k-1,2k+1))$$

Set $\rho^* = \sup_{x\in\mathbb{R}}\rho_1(x)$. We now show for every $\epsilon > 0$ there exists N such that for $k > N$ we have $\mu([k,k+2)) < \frac{\epsilon}{n\rho^*}$. If $b > N$, then $[b-\frac{1}{n}, b+\frac{1}{n}) \subset [k', k'+2)$ for some $k' > N$. Set $k = k'+1$, then $\mu([b-\frac{1}{n}, b+\frac{1}{n})) \leq \mu([k',k'+2)) < \frac{\epsilon}{n\rho^*}$.

Note that since $\rho^* = \sup_{x\in\mathbb{R}}\rho_1(x)$, then $n\rho_1(nx) \leq n\rho^*$, and so

$$\begin{aligned}
\rho_n * \mu(b) :=& \int_{\mathbb{R}} \rho_n(b-y)d\mu(y) \\
=& \int_{b-\frac{1}{n}}^{b+\frac{1}{n}} \rho_n(b-y)d\mu(y) \\
\leq& n\rho^*\mu([b-\frac{1}{n}, b+\frac{1}{n}]) \leq n\rho^*\frac{\epsilon}{n\rho^*} < \epsilon
\end{aligned}$$

□

We can now prove the main theorem of this chapter.

Lemma 3.15. *If μ, ν are finite measures with $\mathcal{F}(\mu) = \mathcal{F}(\nu)$, then $\mu = \nu$.*

Proof. Let μ, ν be as above. We first approximate μ, ν be smooth functions $\rho_n * \mu, \rho_n * \nu$. We have:

$$\mathcal{F}\mu_n(u) = \widehat{\rho_n}(u)\mathcal{F}\mu(u) = \rho_n(u)\mathcal{F}\nu(u) = \mathcal{F}\nu(u) \quad \text{(by Lemma (3.12))}$$

We know that $\rho_n * \mu$ ($\rho_n * \nu$) is continuous by Lemma (3.13). The Fourier transform of these exists by the above computation. We need to know that the Fourier transform is in L^1.

$$\begin{aligned}
|\mathcal{F}(\rho_n * \mu)(u)| &= |\widehat{\rho_n}(u)| \cdot |\mathcal{F}\mu(u)| \\
&= |\int e^{ixu}\rho_n(x)dx| \cdot |\int e^{iux}d\mu(x)| \\
&\leq \int \rho_n(x)|e^{ixu}|dx \int |e^{iux}|d\mu(x) = 1 \cdot \int d\mu(x) \in L^\infty
\end{aligned}$$

Furthermore

$$\begin{aligned}
|\mathcal{F}(\rho_n * \mu)(u)| &= \left| \lim_{a,b\to\infty} \int_{-a}^{b} e^{iux}\rho_n * \mu(x)dx \right| \\
&= \left| \lim_{a,b\to\infty} \left(\frac{1}{iu}e^{iub}(\rho_n * \mu)(b) - \frac{1}{iu}e^{-iua}(\rho_n * \mu)(-a) - \int_{-a}^{b} \frac{1}{iu}e^{iux}\frac{d}{dx}(\rho_n * \mu)(x)dx \right) \right| \\
&\quad \text{(by parts)} \\
&\leq \frac{1}{u}\left| \int_{\mathbb{R}} e^{iux}\frac{d}{dx}(\rho_n * \mu)(x)dx \right| \quad \text{(by Lemma (3.13))} \\
&= \frac{1}{u^2}\left| \int_{\mathbb{R}} e^{iux}\frac{d^2}{dx^2}(\rho_n * \mu)(x)dx \right| \quad \text{(by parts and Lemma (3.13))} \\
&\leq \frac{1}{u^2}\int_{\mathbb{R}} \left| \frac{d^2}{dx^2}(\rho_n * \mu)(x) \right| dx \\
&\leq \frac{1}{u^2}\int_{\mathbb{R}}\int_{\mathbb{R}} \rho_n''(x-y)d\mu(y)dx \\
&\leq \frac{1}{u^2}\int_{\mathbb{R}}\int_{\mathbb{R}} \rho_n''(x-y)dxd\mu(y) \quad \text{(by Tonelli)} \\
&\leq \frac{1}{u^2}
\end{aligned}$$

Thus (Figure (3.2))

$$|\mathcal{F}(\rho_n * \mu)(u)| \leq \begin{cases} c_1, & |u| \leq 1 \\ c_2 u^{-2}, & |u| \geq 1 \end{cases}$$

but the right hand side is in L^1. Thus $\mathcal{F}(\rho_n * \mu)(u) \in L^1$. Similarly $\mathcal{F}(\rho_n * \nu)(u) \in L^1$. Now

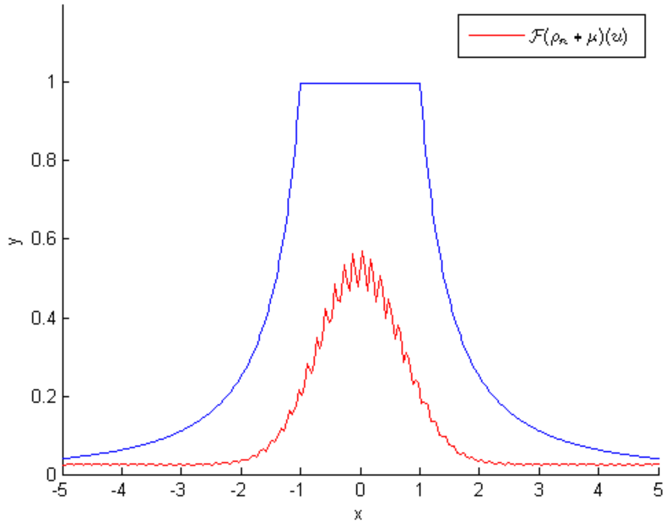

Figure 3.2: *Bound for* $|\mathcal{F}(\rho_n * \mu)(u)|$

by Fourier inversion theorem for functions (Theorem (3.4)):

$$\begin{aligned}\rho_n * \mu(x) =& \frac{1}{2\pi}\int e^{-iux}\mathcal{F}(\rho_n * \mu)(u)du \\ =& \frac{1}{2\pi}\int e^{-iux}\mathcal{F}(\rho_n * \nu)(u)du \quad \text{(by assertion } \mathcal{F}(\mu) = \mathcal{F}(\nu)\text{)} \\ =& \mathcal{F}\rho_n * \nu\end{aligned}$$

Now given $a < b \in \mathbb{R}$ take $a_n \searrow a, b_n \nearrow b$. Define $f(x)$ in a following manner:

$$f_n(x) = \begin{cases} 0, & x < a \\ \frac{1}{a_n - a}(x-a), & a \le x < a_n \\ 1, & a_n \le x < b_n \\ \frac{1}{b_n - b}(x-b), & b_n \le x < b \\ 0, & x \le b \end{cases}$$

Note that $f_n(x)$ is continuous and bounded (Figure (3.3)). Now

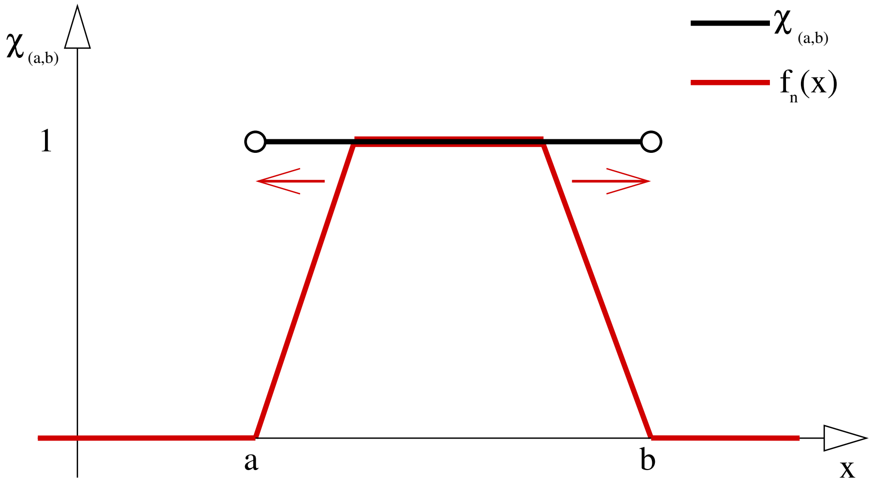

Figure 3.3: $f_n(x)$

$$
\begin{aligned}
\mu((a,b)) &= \int \chi_{(a,b)}(x) d\mu(x) \\
&= \int \lim_{n\to\infty} f_n(x) d\mu(x) \\
&= \lim_{n\to\infty} \int f_n(x) d\mu(x) \quad \text{(by MCT)} \\
&= \lim_{n\to\infty} \lim_{m\to\infty} \int f_n(x) \rho_m * \mu(x) dx \quad (\text{as } \rho_m * \mu \to \mu \text{ narrowly}) \\
&= \lim_{n\to\infty} \lim_{m\to\infty} \int f_n(x) \rho_m * \nu(x) dx \\
&= \lim_{n\to\infty} \int f_n(x) d\nu(x) \\
&= \int \lim_{n\to\infty} f_n(x) d\nu(x) \quad \text{(by MCT)} \\
&= \int \chi_{(a,b)}(x) d\nu(x) = \nu((a,b))
\end{aligned}
$$

□

Lemma 3.16. *Let $(\mu_n)_{n\geq 1}$ be a sequence of probability measures, and let $(\widehat{\mu_n})_{n\geq 1}$ denote their Fourier transforms. If $\widehat{\mu_n}(u)$ converges to a function $f(u)$ for all u and $f(u)$ is continuous at 0, then there exists a probability measure μ such that μ_n converges narrowly to μ and $f(u) = \widehat{\mu}(u)$.*[23]

Proof. Suppose that $\lim_{n\to\infty} \widehat{\mu}(u) = f(u)$.

$$\begin{aligned}
\int_{-\alpha}^{\alpha} d\widehat{\mu_n}(u) &= \int_{-\alpha}^{\alpha} \left(\int_{-\infty}^{\infty} e^{iux} d\mu_n(x) \right) du \\
&= \int_{-\infty}^{\infty} \left(\int_{-\alpha}^{\alpha} e^{iux} du \right) d\mu_n(x) \quad \text{(by Fubini)} \\
&= \int_{-\infty}^{\infty} \left(\int_{-\alpha}^{\alpha} \cos(ux) + i \sin(ux) du \right) d\mu_n(x) \quad (e^{iux} = \cos(ux) + i\sin(ux)) \\
&= \int_{-\infty}^{\infty} \frac{2}{x} \sin(\alpha x) d\mu_n(x) \quad \text{(sin is odd, so integral is zero over symmetric interval)}
\end{aligned}$$

Now we rely on the previous computation to estimate the following integral:

$$\begin{aligned}
\frac{1}{\alpha} \int_{-\alpha}^{\alpha} (1 - \widehat{\mu}(u)) du =& 2 - \int_{-\infty}^{\infty} \frac{2}{\alpha x} \sin(\alpha x) d\mu_n(x) \quad (\int_{-\alpha}^{\alpha} 1 du = 2\alpha) \\
=& 2 \int_{-\infty}^{\infty} \left(1 - \frac{\sin(\alpha x)}{\alpha x} \right) d\mu_n(x) \quad (\int_{-\infty}^{\infty} 1 d\mu_n(x) = 1)
\end{aligned}$$

Note that $2(2 - \frac{\sin v}{v}) \geq 1$ for $|v| \geq 2$ and $2(2 - \frac{\sin v}{v}) \geq 0$ for any v. Using this inequalities, we can estimate the previous interval:

$$\begin{aligned}
2 \int_{-\infty}^{\infty} \left(1 - \frac{\sin(\alpha x)}{\alpha x} \right) d\mu_n(x) \geq& \int_{-\infty}^{\infty} 1_{[-2,2]^c}(\alpha x) d\mu_n(x) \\
=& \int 1_{[-\frac{2}{\alpha}, \frac{2}{\alpha}]^c}(x) d\mu_n(x) \\
=& \mu_n \left(\left[-\frac{2}{\alpha}, \frac{2}{\alpha} \right]^c \right)
\end{aligned}$$

Let $\beta = \frac{2}{\alpha}$, so the estimate reads as

$$\mu_n([-\beta, \beta]^c) \leq \frac{\beta}{2} \int_{-\frac{2}{\beta}}^{\frac{2}{\beta}} (1 - \widehat{\mu_n}(u)) du \tag{3.5}$$

Note that $\widehat{\mu_n}(0) = 1$ for all n, whence $\lim_{n\to\infty} \widehat{\mu_n}(0) = f(0) = 1$ by initial assumption. Now let $\epsilon > 0$, then there exists $\alpha > 0$ such that $|f(0) - f(u)| = |1 - f(u)| < \frac{\epsilon}{4}$ if $|u| < \frac{2}{\alpha}$ by continuity of f at 0. Thus,

$$\left| \frac{\alpha}{2} \int_{-\frac{2}{\alpha}}^{\frac{2}{\alpha}} (1 - f(u)) du \right| \leq \frac{\alpha}{2} \int_{-\frac{2}{\alpha}}^{\frac{2}{\alpha}} \frac{\epsilon}{4} du = \frac{\epsilon}{2} \tag{3.6}$$

Since $|\widehat{\mu_n}(u)| = |\int e^{iux} d\mu_n(x)| \leq \int |e^{iux}| d\mu_n(x) \leq \int 1 d\mu_n(x) = 1$, so by Lebesgue's dominated convergence theorem

$$\lim_{n\to\infty} \frac{\alpha}{2} \int_{-\frac{2}{\alpha}}^{\frac{2}{\alpha}} (1 - \widehat{\mu_n}(u)) du = \frac{\alpha}{2} \int_{-\frac{2}{\alpha}}^{\frac{2}{\alpha}} (1 - \lim_{n\to\infty} \widehat{\mu_n}(u)) du = \frac{\alpha}{2} \int_{-\frac{2}{\alpha}}^{\frac{2}{\alpha}} (1 - f(u)) du$$

So there exists N such that for all $n \geq N$

$$\left| \int_{-\frac{2}{\alpha}}^{\frac{2}{\alpha}} (1 - \widehat{\mu_n}(u)) du - \int_{-\frac{2}{\alpha}}^{\frac{2}{\alpha}} (1 - f(u)) du \right| \leq \frac{\epsilon}{\alpha}$$

Once one multiplies the inequality above by $\frac{\alpha}{2}$, applies triangle inequality as well as (3.6) one gets $\frac{\alpha}{2} \int_{-\frac{2}{\alpha}}^{\frac{2}{\alpha}} (1 - \widehat{\mu_n}(u)) du \leq \epsilon$. We next use estimate (3.5) to conclude $\mu_n([-\alpha, \alpha]^c) \leq \epsilon$ for all $n > N$.

So for each $n \leq N$, there exists α_n such that $\mu_n([-\alpha_n, \alpha_n]^c) \leq \epsilon$. Let $a = \max(\alpha_1, \ldots, \alpha_n; \alpha)$ (N is finite). Then we have $\mu_n([-a, a]^c) \leq \epsilon$ for all n. In other words for the sequence $(\mu_n)_{n \geq 1}$, for any $\epsilon > 0$ there exists $a \in \mathbb{R}$ such that $\sup_n \mu_n([-a, a]^c) \leq \epsilon$. Therefore we have shown

$$\limsup_{m \to \infty} \sup_n \mu_n([-m, m]^c) = 0.$$

We can next apply theorem (3.10) to obtain a subsequence $(n_k)_{k \geq 1}$ such that μ_{n_k} converges narrowly to μ as $k \to \infty$. By the same theorem $\lim_{k \to \infty} \widehat{\mu_{n_k}}(u) = \widehat{\mu}(u)$ for all u, thus $f(u) = \widehat{\mu}(u)$.

We now show that the sequence $(\mu_n)_{n \geq 1}$ converges narrowly to μ by the method of contradiction. Let F_n, F be distribution functions of μ_n and μ, $D = \{x : F(-x) = F(x)\}$. Suppose that μ_n does not converge narrowly to μ, then by Lemma (3.9) there must be at least one point $x \in D$ and a subsequence $(n_k)_{k \geq 1}$ such that $\lim_{k \to \infty} F_{n_k}(x)$ exists and moreover $\lim_{k \to \infty} F_{n_k}(x) = \beta \neq F(x)$. Next by theorem (3.10) there also exists a subsequence of the sequence (n_k) (denote it as $(n_{k_j})_{j \geq 1}$), such that $(\mu_{n_{k_j}})_{j \geq 1}$ converges weakly to a limit ν as $j \to \infty$. Thus we get

$$\lim_{j \to \infty} \widehat{\mu_{n_{k_j}}}(u) = \widehat{\nu}(u)$$

and since $\lim \widehat{\mu_n}(u) = f(u)$, we conclude $\widehat{\nu}(u) = f(u)$. But we have seen that $f(u) = \widehat{\mu}(u)$, therefore by Lemma (3.15) we must have $\mu = \nu$. Since $\mu_{n_{k_j}}(u)$ converges to $\mu = \nu$ by Lemma (3.9) and $x \in D$, we obtain $\lim_{j \to \infty} F_{n_{k_j}}(x) = F(x)$. But $\lim_{j \to \infty} F_{n_{k_j}}(x) = \beta \neq F(x)$, and we have a contradiction. $\square$

Chapter 4

Central Limit Theorem

Definition 4.1. *Let $(X_n)_{n\geq 1}, X$ be random variables. We say X_n converges in distribution (law) to X if the distribution measures P^{X_n} converge narrowly to P^X.*

Lemma 4.2.

$$\left| n \left[e^{iu\frac{x-\mu}{\sigma\sqrt{n}}} - (1 + iu\frac{x-\mu}{\sigma\sqrt{n}} - \frac{1}{2}u^2\frac{(x-\mu)^2}{\sigma^2 n}) \right] \right| \leq \frac{(x-\mu)^2}{\sigma^2}u^2 \in L^1(\mathbb{P}_{X_j})$$

Proof. Consider the Taylor expansion of the following function: $f(u) = e^{iu\frac{x-\mu}{\sigma\sqrt{n}}}$.

$$\begin{aligned}
&f'(u) = i\frac{x-\mu}{\sigma\sqrt{n}}e^{iu\frac{x-\mu}{\sigma\sqrt{n}}} \\
&f''(u) = -\frac{(x-\mu)^2}{\sigma^2 n}e^{iu\frac{x-\mu}{\sigma\sqrt{n}}} \\
&f(u) = e^{iu\frac{x-\mu}{\sigma\sqrt{n}}} = f(0) + f'(0)u + \frac{1}{2!}f''(c)u^2 = \qquad (-|u| < c < |u|) \\
=&1 + iu\frac{x-\mu}{\sigma\sqrt{n}} - \frac{1}{2}\frac{(x-\mu)^2}{\sigma^2 n}e^{ic\frac{x-\mu}{\sigma\sqrt{n}}}
\end{aligned}$$

So

$$\begin{aligned}
&\left| n \left[e^{iu\frac{x-\mu}{\sigma\sqrt{n}}} - (1 + iu\frac{x-\mu}{\sigma\sqrt{n}} - \frac{1}{2}u^2\frac{(x-\mu)^2}{\sigma^2 n}) \right] \right| = \\
=&\left| n \left[e^{iu\frac{x-\mu}{\sigma\sqrt{n}}} - (1 + iu\frac{x-\mu}{\sigma\sqrt{n}}) \right] - \frac{1}{2}u^2\frac{(x-\mu)^2}{\sigma^2}) \right| \leq \\
\leq&\left| n \left[e^{iu\frac{x-\mu}{\sigma\sqrt{n}}} - (1 + iu\frac{x-\mu}{\sigma\sqrt{n}}) \right] \right| + \frac{1}{2}u^2\frac{(x-\mu)^2}{\sigma^2} \leq \\
\leq&\frac{(x-\mu)^2}{2\sigma^2}u^2\left| e^{ic\frac{x-\mu}{\sigma\sqrt{n}}} \right| + \frac{1}{2}u^2\frac{(x-\mu)^2}{\sigma^2} \leq \\
\leq&\frac{(x-\mu)^2}{\sigma^2}u^2 \in L^1(\mathbb{P}_{X_j})
\end{aligned}$$

□

Theorem 4.3. *Let $\{X_j\}_{j\geq 1}$ be a sequence of independent and identically distributed random variables with $\mathbb{E}(X_j) = \mu$ and $Var(X_j) = \sigma^2$ with $0 < \sigma < \infty$. Let $S_n = \sum_{j=1}^n X_j$, $Y_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$. Then Y_n converges in distribution to the normal distribution with mean 0 and variance 1.*

Proof. Let $\{X_j\}_{j\geq 1}, S_n, Y_n$ be as above. Consider

$$\begin{aligned}
\mathbb{E}(e^{iuY_n}) &= \mathbb{E}(e^{iu\frac{\sum_{j=1}^n (X_j-\mu)}{\sigma\sqrt{n}}}) \\
&= \mathbb{E}(\prod_{j=1}^n e^{iu\frac{X_j-\mu}{\sigma\sqrt{n}}}) \\
&= \prod_{j=1}^n \mathbb{E}(e^{iu\frac{X_j-\mu}{\sigma\sqrt{n}}}) \\
&= \prod_{j=1}^n \mathbb{E}(1 + iu\frac{X_j-\mu}{\sigma\sqrt{n}} - \frac{1}{2}u^2\frac{(X_j-\mu)^2}{\sigma^2 n} + o(\frac{u^2}{n})) \quad \text{(Taylor)} \\
&= \prod_{j=1}^n (\mathbb{E}(1) + \mathbb{E}(iu\frac{X_j-\mu}{\sigma\sqrt{n}}) - \mathbb{E}(\frac{1}{2}u^2\frac{(X_j-\mu)^2}{\sigma^2 n}) + \epsilon(n))) \\
&\quad \text{where } \epsilon(n) = \int_{\mathbb{R}} \left[e^{iu\frac{x-\mu}{\sigma\sqrt{n}}} - (1 + iu\frac{x-\mu}{\sigma\sqrt{n}} - \frac{1}{2}u^2\frac{(x-\mu)^2}{\sigma^2 n}) \right] d\,\mathbb{P}_{X_j}(x) \\
&= \prod_{j=1}^n (1 + iu\frac{\mathbb{E}(X_j-\mu)}{\sigma\sqrt{n}} - \frac{1}{2}u^2\frac{\mathbb{E}((X_j-\mu)^2)}{\sigma^2 n} + \epsilon(n)) \quad (\mathbb{E} \text{ is linear}) \\
&= \prod_{j=1}^n (1 - \frac{1}{2}u^2\frac{\mathbb{E}((X_j-\mu)^2)}{\sigma^2 n} + \epsilon(n)) \quad (\mathbb{E}(X_j-\mu) = 0) \\
&= \prod_{j=1}^n (1 - \frac{u^2}{2n} + \epsilon(n)) \quad (\mathbb{E}((X_j-\mu)^2) = \sigma^2)
\end{aligned}$$

Now we will show that $\lim_{n\to\infty} n\epsilon(n) = 0$. Note that Lemma (4.2) will allow us to use Lebesgue's dominated convergence theorem.

$$\begin{aligned}
\lim_{n\to\infty} n\epsilon(n) &= \lim_{n\to\infty} \int_{\mathbb{R}} n \left[e^{iu\frac{x-\mu}{\sigma\sqrt{n}}} - (1 + iu\frac{x-\mu}{\sigma\sqrt{n}} - \frac{1}{2}u^2\frac{(x-\mu)^2}{\sigma^2 n}) \right] d\mathbb{P}_{X_j}(x) \\
&= \int_{\mathbb{R}} \lim_{n\to\infty} n \left[e^{iu\frac{x-\mu}{\sigma\sqrt{n}}} - (1 + iu\frac{x-\mu}{\sigma\sqrt{n}} - \frac{1}{2}u^2\frac{(x-\mu)^2}{\sigma^2 n}) \right] d\mathbb{P}_{X_j}(x) \quad \text{(DCT)} \\
&= \int_{\mathbb{R}} 0 \, d\mathbb{P}_{X_j}(x) = 0 \quad \text{(Taylor)}
\end{aligned}$$

Now let us take a look at the limit:

$$\begin{aligned}
&\lim_{n\to\infty}\prod_{j=1}^{n}(1-\frac{u^2}{2n}+\epsilon(n)) = \\
&= \lim_{n\to\infty}(1-\frac{u^2}{2n}+\epsilon(n))^n = \\
&= \lim_{n\to\infty} e^{n\ln(1-\frac{u^2}{2n}+\epsilon(n))} = \\
&= \lim_{n\to\infty} e^{-\frac{1}{2}u^2+n\epsilon(n)+no(\frac{u^2}{2n}+\epsilon(n))} = \qquad \text{(Taylor of } ln(1-x))\\
&=e^{\lim_{n\to\infty}\left(-\frac{1}{2}u^2+n\epsilon(n)+no(\frac{u^2}{2n}+\epsilon(n))\right)} = \qquad (e^x \text{ is continuous})\\
&=e^{-\frac{u^2}{2}}
\end{aligned}$$

As $e^{-\frac{u^2}{2}}$ is continuous, Lemma (3.16) proves that there is a measure μ with $\mathbb{P}^{Y_n} \to \mu$ such that $\mathcal{F}(\mathbb{P}_{Y_n}) \to \mathcal{F}(\mu)$ and $\mathcal{F}(\mathbb{P}_{Y_n}) \to e^{-\frac{u^2}{2}} = \mathcal{F}(N)$, so $\mathcal{F}(\mu) = \mathcal{F}(N)$, thus $\mu = N$ by Lemma (3.15). $\square$

Bibliography

[1] R. Bass, Stochastic calculus, Website, 2001, http://www.math.uconn.edu/~bass/stochcalc.pdf.

[2] J. Jacod and P. Protter, *Probability essentials*, Universitext, Springer-Verlag, Berlin, second edition, 2003.

[3] E. M. Stein and R. Shakarchi, *Fourier analysis*, volume 1 of *Princeton Lectures in Analysis*, Princeton University Press, Princeton, NJ, 2003, An introduction.