LECTURE NOTES

MEASURE THEORY and PROBABILITY

Rodrigo Bañuelos
Department of Mathematics
Purdue University
West Lafayette, IN 47907

June 20, 2003

I

SIGMA ALGEBRAS AND MEASURES

§1 $\sigma-$Algebras: Definitions and Notation.

We use Ω to denote an abstract space. That is, a collection of objects called points. These points are denoted by ω. We use the standard notation: For $A, B \subset \Omega$, we denote $A \cup B$ their union, $A \cap B$ their intersection, A^c the complement of A, $A \backslash B = A - B = \{x \in A : x \notin B\} = A \cap B^c$ and $A \Delta B = (A \backslash B) \cup (B \backslash A)$. If $A_1 \subset A_2, \ldots$ and $A = \cup_{n=1}^{\infty} A_n$, we will write $A_n \uparrow A$. If $A_1 \supset A_2 \supset \ldots$ and $A = \cap_{n=1}^{\infty} A_n$, we will write $A_n \downarrow A$. Recall that $(\cup_n A_n)^c = \cap_n A_n^c$ and $(\cap_n A_n)^c = \cup_n A_n^c$. With this notation we see that $A_n \uparrow A \Rightarrow A_n^c \downarrow A^c$ and $A_n \downarrow A \Rightarrow A_n^c \uparrow A^c$. If $A_1, \ldots, A_n \in \Omega$, we can write

$$\cup_{j=1}^{n} A_j = A_1 \cup (A_1^c \cap A_2) \cup (A_1^c \cap A_2^c \cap A_3) \cup \ldots (A_1^c \cap \ldots \cap A_{n-1}^c \cap A_n), \quad (1.1)$$

which is a disjoint union of sets. In fact, this can be done for infinitely many sets:

$$\cup_{n=1}^{\infty} A_n = \cup_{n=1}^{\infty} (A_1^c \cap \ldots \cap A_{n-1}^c \cap A_n). \quad (1.2)$$

If $A_n \uparrow$, then

$$\cup_{j=1}^{n} A_j = A_1 \cup (A_2 \backslash A_1) \cup (A_3 \backslash A_2) \ldots \cup (A_n \backslash A_{n-1}). \quad (1.3)$$

Two sets which play an important role in studying convergence questions are:

$$\overline{\lim} A_n) = \limsup_n A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \quad (1.4)$$

and

$$\underline{\lim} A_n = \liminf_n A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k. \quad (1.5)$$

Notice

$$
\begin{aligned}
(\overline{\lim} A_n)^c &= \left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_n \right)^c \\
&= \bigcup_{n=1}^{\infty} \left(\bigcup_{k=n}^{\infty} A_k \right)^c \\
&= \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c = \underline{\lim} A_n^c
\end{aligned}
$$

Also, $x \in \overline{\lim} A_n$ if and only if $x \in \bigcup_{k=n}^{\infty} A_k$ for all n. Equivalently, for all n there is at least one $k > n$ such that $x \in A_{k_0}$. That is, $x \in A_n$ for infinitely many n. For this reason when $x \in \overline{\lim} A_n$ we say that x belongs to infinitely many of the $A_n' s$ and write this as $x \in A_n$ $i.o.$ If $x \in \underline{\lim} A_n$ this means that $x \in \bigcap_{k=n}^{\infty} A_k$ for some n or equivalently, $x \in A_k$ for all $k > n$. For this reason when $x \in \underline{\lim} A_n$ we say that $x \in A_n$, *eventually*. We will see connections to $\overline{\lim} x_k$, $\underline{\lim} x_k$, where $\{x_k\}$ is a sequence of points later.

Definition 1.1. Let $\mathcal{F}$ be a collection of subsets of Ω. $\mathcal{F}$ is called a field (algebra) if $\Omega \in \mathcal{F}$ and $\mathcal{F}$ is closed under complementation and finite union. That is,

(i) $\Omega \in \mathcal{F}$

(ii) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$

(ii) $A_1, A_2, \dots A_n \in \mathcal{F} \Rightarrow \bigcup_{j=1}^{n} A_j \in \mathcal{F}$.

If in addition, (iii) can be replaced by countable unions, that is if

(iv) $A_1, \dots A_n, \dots \in \mathcal{F} \Rightarrow \bigcup_{j=1}^{\infty} A_j \in \mathcal{F}$,
then $\mathcal{F}$ is called a σ*–algebra* or often also a σ*–field.*

Here are three simple examples of σ*–algebras.*

(i) $\mathcal{F} = \{\emptyset, \Omega\}$,

(ii) $\mathcal{F} = \{\text{all subsets of } \Omega\}$,

(iii) If $A \subset \Omega$, $\mathcal{F} = \{\emptyset, \Omega, A, A^c\}$.

An example of an algebra which is not a σ–algebra is given by the following. Let $\Omega = \mathbb{R}$, the real numbers and take $\mathcal{F}$ to be the collection of all finite disjoint unions of intervals of the form $(a, b] = \{x: a < x \leq b\}$, $-\infty \leq a < b < \infty$. By convention we also count (a, ∞) as right–semiclosed. $\mathcal{F}$ is an algebra but not a σ–algebra. Set

$$A_n = (0, 1 - \frac{1}{n}].$$

Then,

$$\bigcup_{n=1}^{\infty} A_n = (0, 1) \notin \mathcal{F}.$$

The convention is important here because $(a, b]^c = (b, \infty) \cup (-\infty, a]$.

Remark 1.1. *We will refer to the pair* $(\Omega, \mathcal{F})$ *as a measurable space. The reason for this will become clear in the next section when we introduce measures.*

Definition 1.2. Given any collection $\mathcal{A}$ of subsets of Ω, let $\sigma(\mathcal{A})$ be the smallest σ–algebra containing $\mathcal{A}$. That is if $\mathcal{F}$ is another σ–algebra and $\mathcal{A} \subset \mathcal{F}$, then $\sigma(\mathcal{A}) \subset \mathcal{F}$.

Is there such a σ–algebra? The answer is, of course, yes. In fact,

$$\sigma(\mathcal{A}) = \bigcap \mathcal{F}$$

where the intersection is take over all the σ–algebras containing the collection $\mathcal{A}$. This collection is not empty since $\mathcal{A} \subset$ all subsets of Ω which is a σ–algebra. We call $\sigma(\mathcal{A})$ the σ–algebra generated by $\mathcal{A}$. If $\mathcal{F}_0$ is an algebra, we often write $\sigma(\mathcal{F}_0) = \overline{\mathcal{F}}_0$.

Example 1.1. $\mathcal{A} = \{A\}$, $A \subset \Omega$. *Then*

$$\sigma(\mathcal{A}) = \{\emptyset, A, A^c, \Omega\}.$$

Problem 1.1. *Let $\mathcal{A}$ be a collection of subsets of Ω and $A \subset \Omega$. Set $\mathcal{A} \cap A = \{B \cap A \colon B \in \mathcal{A}\}$. Assume $\sigma(\mathcal{A}) = \mathcal{F}$. Show that $\sigma(\mathcal{A} \cap A) = \mathcal{F} \cap A$, relative to A.*

Definition 1.2. Let $\Omega = \mathbb{R}$ and $\mathcal{B}_0$ the field of right–semiclosed intervals. Then $\sigma(\mathcal{B}_0) = \mathcal{B}$ is called the Borel σ–algebra of $\mathbb{R}$.

Problem 1.2. *Prove that every open set in $\mathbb{R}$ is the countable union of right –semiclosed intervals.*

Problem 1.3. *Prove that every open set is in $\mathcal{B}$.*

Problem 1.4. *Prove that $\mathcal{B} = \sigma(\{$all open intervals$\})$.*

Remark 1.2. *The above construction works equally in $\mathbb{R}^d$ where we take $\mathcal{B}_0$ to be the family of all intervals of the form*

$$(a_1, b_1] \times \dots (a_d, b_d], \qquad -\infty \le a_i < b_i < \infty.$$

§2. Measures.

Definition 2.1. *Let $(\Omega, \mathcal{F})$ be a measurable space. By a measure on this space we mean a function $\mu : \mathcal{F} \to [0, \infty]$ with the properties*

(i) $\mu(\emptyset) = 0$

and

(ii) if $A_j \in \mathcal{F}$ are disjoint then

$$\mu\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mu(A_j).$$

Remark 2.1. *We will refer to the triple* $(\Omega, \mathcal{F}, \mu)$ *as a measure space. If* $\mu(\Omega) = 1$ *we refer to it as a probability space and often write this as* $(\Omega, \mathcal{F}, P)$.

Example 2.1. Let Ω be a countable set and let $\mathcal{F}$ = collection of all subsets of Ω. Denote by $\#A$ denote the number of point in A. Define $\mu(A) = \#A$. This is called the counting measure. If Ω is a finite set with n points and we define $P(A) = \frac{1}{n}\#A$ then we get a probability measure. Concrete examples of these are:

(i) Coin flips. Let $\Omega = \{0, 1\} = \{\text{Heads, Tails}\} = \{T, H\}$ and set $P\{0\} = 1/2$ and $P\{1\} = 1/2$

(2) Rolling a die. $\Omega = \{1, 2, 3, 4, 5, 6\}$, $P\{w\} = 1/6$.

Of course, these are nothing but two very simple examples of probability spaces and our goal now is to enlarge this collection. First, we list several elementary properties of general measures.

Proposition 2.1. *Let* $(\Omega, \mathcal{F}, \mu)$ *be a measure space. Assume all sets mentioned below are in* $\mathcal{F}$.

(i) If $A \subset B$, *then* $\mu(A) \leq \mu(B)$, *(monotonicity).*

(ii) If $A \subseteq \bigcup_{j=1}^{\infty} A_j$, *then* $\mu(A) \leq \sum_{j=1}^{\infty} \mu(A_j)$, *(subadditivity).*

(iii) If $A_j \uparrow A$, *then* $\mu(A_j) \uparrow \mu(A)$, *(continuity for below).*

(iv) If $A_j \downarrow A$ *and* $\mu(A_1) < \infty$, *then* $\mu(A_j) \downarrow \mu(A)$, *(continuity from above).*

Remark 2.2. *The finiteness assumption in (iv) is needed. To see this, set* $\Omega = \{1, 2, 3, \ldots\}$ *and let* μ *be the counting measure. Let* $A_j = \{j,\ j+1, \ldots\}$. *Then* $A_j \downarrow \emptyset$ *but* $\mu(A_j) = \infty$ *for all* j.

Proof. Write $B = A \cup (B \backslash A)$. Then

$$\mu(B) = \mu(A) + \mu(B \backslash A) \geq \mu(A),$$

which proves (i). As a side remark here, note that if if $\mu(A) < \infty$, we have $\mu(B\backslash A) = \mu(B) - \mu(A)$. Next, recall that $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} (A_1^c \cap \ldots \cap A_{n-1}^c \cap A_n)$ where the sets in the last union are disjoint. Therefore,

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n \cap A_1^c \ldots \cap A_{n-1}^c) \leq \sum_{n=1}^{\infty} \mu(A_n),$$

proving (ii).

For (iii) observe that if $A_n \uparrow$, then

$$\begin{aligned}
\mu\left(\bigcup_{n=1}^{\infty} A_n\right) &= \mu\left(\bigcup_{n=1}^{\infty} (A_n \backslash A_{n-1})\right) \\
&= \sum_{n=1}^{\infty} \mu(A_n \backslash A_{n-1}) \\
&= \lim_{m\to\infty} \sum_{n=1}^{m} \mu(A_n \backslash A_{n-1}) \\
&= \lim_{m\to\infty} \mu\left(\bigcup_{n=1}^{m} A_n \backslash A_{n-1}\right).
\end{aligned}$$

For (iv) we observe that if $A_n \downarrow A$ then $A_1 \backslash A_n \uparrow A_1 \backslash A$. By (iii), $\mu(A_1 \backslash A_n) \uparrow \mu(A_1 \backslash A)$ and since $\mu(A_1 \backslash A_n) = \mu(A_1) - \mu(A_n)$ we see that $\mu(A_1) - \mu(A_n) \uparrow \mu(A_1) - \mu(A)$, from which the result follows assuming the finiteness of $\mu(A_1)$.

Definition 2.2. A Lebesgue–Stieltjes measure on $\mathbb{R}$ is a measure on $\mathcal{B} = \sigma(\mathcal{B}_0)$ such that $\mu(I) < \infty$ for each bounded interval I. By an extended distribution function on $\mathbb{R}$ we shall mean a map $F\colon \mathbb{R} \to \mathbb{R}$ that is increasing, $F(a) \leq F(b)$ if $a < b$, and right continuous, $\lim_{x\to x_0^+} F(x) = F(x_0)$. If in addition the function F is nonnegative satisfying $\lim_{x\to\infty} F_X(x) = 1$ and $\lim_{x\to-\infty} F_X(x) = 0$, we shall simply call it a distribution function.

We will show that the formula $\mu(a,b] = F(b) - F(a)$ sets a 1-1 correspondence between the Lebesgue–Stieltjes measures and distributions where two distributions that differ by a constant are identified. of course, probability measures correspond to distributions.

Proposition 2.2. *Let μ be a Lebesgue–Stieltjes measure on $\mathbb{R}$. Define $F\colon \mathbb{R} \to \mathbb{R}$, up to additive constants, by $F(b) - F(a) = \mu(a, b]$. For example, fix $F(0)$ arbitrary and set $F(x) - F(0) = \mu(0, x],\ x \geq 0,\ F(0) - F(x) = \mu(x, 0],\ x < 0$. Then F is an extended distribution.*

Proof. Let $a < b$. Then $F(b) - F(a) = \mu(a, b] \geq 0$. Also, if $\{x_n\}$ is such that $x_1 > x_2 > \ldots \to x$, then $\mu(x, x_n] \to 0$, by Proposition 2.1, (iv), since $\bigcap\limits_{n=1}^{\infty}(x_1, x_n] = \emptyset$ and $(x, x_n] \downarrow \emptyset$. Thus $F(x_n) - F(x) \to 0$ implying that F is right continuous.

We should notice also that

$$\begin{aligned}\mu\{b\} &= \lim_{n\to\infty} \mu\left(b - \frac{1}{n},\ b\right]\\ &= \lim_{n\to\infty} F(b) - F(b - 1/n) = F(b) - F(b-).\end{aligned}$$

Hence in fact F is continued at $\{b\}$ if and only if $\mu\{b\} = 0$.

Problem 2.1. *Set $F(x-) = \lim\limits_{x\to x^-} F(x)$. Then*

$$\mu(a, b) = F(b^-) - F(a) \tag{1}$$

$$\mu[a, b] = F(b) - F(a^-) \tag{2}$$

$$\mu[a, b) = F(b^-) - F(a^-) \tag{3}$$

$$\mu(\mathbb{R}) = F(\infty) - F(-\infty). \tag{4}$$

Theorem 2.1. *Suppose F is a distribution function on $\mathbb{R}$. There is a unique measure μ on $\mathcal{B}(\mathbb{R})$ such that $\mu(a, b] = F(b) - F(a)$.*

Definition 2.3. *Suppose $\mathcal{A}$ is an algebra. μ is a measure on $\mathcal{A}$ if $\mu : \mathcal{A} \to [0, \infty]$, $\mu(\emptyset) = 0$ and if $A_1, A_2, \ldots$ are disjoint with $A = \bigcup_j^{\infty} A_j \in \mathcal{A}$, then $\mu(A) = \sum_j^{\infty} \mu(A_j)$. The measure is σ–finite if the space $\Omega = \cup_{j=1}^{\infty} \Omega_j$ where the $\Omega_j \in \mathcal{A}$ are disjoint and $\mu(\Omega_j) < \infty$.*

Theorem 2.2 (Caratheodory's Extension Theorem). *Suppose μ is σ–finite on an algebra $\mathcal{A}$. Then μ has a unique extension to $\sigma(\mathcal{A})$.*

We return to the proof of this theorem later.

Definition 2.4. *A collection $\mathcal{S} \subset \Omega$ is a semialgebra if the following two conditions hold.*

(i) $A,\ B \in \mathcal{S} \Rightarrow A \cap B \in \mathcal{S}$,

(ii) $A \in \mathcal{S}$ *then* A^c *is the finite union of disjoint sets in* $\mathcal{S}$.

Example 2.2. $\mathcal{S} = \{(a, b]\colon -\infty \le a < b < \infty\}$. *This is a semialgebra but not an algebra.*

Lemma 2.1. *If $\mathcal{S}$ is a semi-algebra, then $\overline{\mathcal{S}} = \{$finite disjoint unions of sets in $\mathcal{S}\}$ is an algebra. This is called the algebra generated by $\mathcal{S}$.*

Proof. Let $E_1 = \cup_{j=1}^n A_j$ and $E_2 = \cup_{j=1}^n B_j$, where the unions are disjoint and the sets are all in $\mathcal{S}$. Then $E_1 \cap E_2 = \cup_{i,j} A_i \cap B_j \in \overline{\mathcal{S}}$. Thus $\overline{\mathcal{S}}$ is closed under finite intersections. Also, if $E = \cup_{j=1}^n A_j \in \overline{\mathcal{S}}$ then $A^c = \cap_j A_j^c$. However, by the definition of S, and $\overline{\mathcal{S}}$, we see that $A^c \in \overline{\mathcal{S}}$. This proves that $\overline{\mathcal{S}}$ is an algebra.

Theorem 2.3. *Let $\mathcal{S}$ be a semialgebra and let μ be defined on $\mathcal{S}$. Suppose $\mu(\emptyset) = 0$ with the additional properties:*

(i) *If* $E \in \mathcal{S},\ E = \bigcup_{i=1}^n E_i,\ E_i \in \mathcal{S}$ *disjoint, then* $\mu(E) = \sum_{i=1}^n \mu(E_i)$

and

(ii) *If* $E \in \mathcal{S},\ E = \bigcup_{i=1}^\infty E_i,\ E_i \in \mathcal{S}$ *disjoint, then* $\mu(E) \le \sum_{i=1}^\infty \mu(E_i)$.

Then μ has a unique extension $\overline{\mu}$ to $\overline{\mathcal{S}}$ which is a measure. In addition, if $\overline{\mu}$ is σ–finite, then $\overline{\mu}$ has a unique extension to a measure (which we continuo to call $\overline{\mu}$) to $\sigma(\mathcal{S})$, by the Carathéodary extension theorem

Proof. Define $\overline{\mu}$ on $\overline{\mathcal{S}}$ by

$$\overline{\mu}(E) = \sum_{i=1}^{n} \mu(E_i),$$

if $E = \bigcup_{i=1}^{n} E_i,\ E_i \in \mathcal{S}$ and the union is disjoint. We first verify that this is well defined. That is, suppose we also have $E = \bigcup_{j=1}^{m} \tilde{E}_j$, where $\tilde{E}_j \in \mathcal{S}$ and disjoint. Then

$$E_i = \bigcup_{j=1}^{m} (E_i \cap \tilde{E}_j), \quad \tilde{E}_j = \bigcup_{i=1}^{n} (E_i \cap \tilde{E}_j).$$

By (i),

$$\sum_{i=1}^{n} \mu(E_i) = \sum_{i=1}^{n} \mu(E_i \cap \tilde{E}_j) = \sum_{m} \sum_{n} \mu(E_i \cap \tilde{E}_j) = \sum_{m} \mu(\tilde{E}_j).$$

So, $\overline{\mu}$ is well defined. It remains to verify that $\overline{\mu}$ so defined is a measure. We postpone the proof of this to state the following lemma which will be used in its proof.

Lemma 2.2. *Suppose (i) above holds and let $\overline{\mu}$ be defined on $\overline{\mathcal{S}}$ as above.*

(a) If $E,\ E_i \in \overline{\mathcal{S}}$, E_i disjoint, with $E = \bigcup\limits_{i=1}^{n} E_i$. Then $\overline{\mu}(E) = \sum\limits_{i=1}^{n} \overline{\mu}(E_i)$.

(b) If $E,\ E_i \in \overline{\mathcal{S}},\ E \subset \bigcup\limits_{i=1}^{n} E_i$, then $\overline{\mu}(E) \leq \sum\limits_{i=1}^{n} \overline{\mu}(E_i)$.

Note that (a) gives more than (i) since $E_i \in \overline{\mathcal{S}}$ not just $\mathcal{S}$. Also, the sets in (b) are not necessarily disjoint. We assume the Lemma for the moment.

Next, let $E = \bigcup\limits_{i=1}^{\infty} E_i,\ E_i \in \overline{\mathcal{S}}$ where the sets are disjoint and assume, as required by the definition of the measures on *algebras*, that $E \in \overline{\mathcal{S}}$. Since $E_i = \bigcup\limits_{j=1}^{n} E_{ij},\ E_{ij} \in \mathcal{S}$, with these also disjoint, we have

$$\sum_{i=1}^{\infty} \overline{\mu}(E_i) \overset{(i)}{=} \sum_{i=1}^{\infty} \sum_{j=1}^{n} \overline{\mu}(E_{ij}) = \sum_{i,j} \mu(E_{ij}).$$

So, we may assume $E_i \in \mathcal{S}$ instead of $\overline{\mathcal{S}}$, otherwise replace it by E_{ij}. Since $E \in \overline{\mathcal{S}},\ E = \bigcup\limits_{j=1}^{n} \tilde{E}_j,\ \tilde{E}_j \in \mathcal{S}$ and again disjoint sets, and we can write

$$\tilde{E}_j = \bigcup_{i=1}^{\infty} (\tilde{E}_j \cap E_i).$$

Thus by assumption (ii),

$$\mu(\tilde{E}_j) \leq \sum_{i=1}^{\infty} \mu(\tilde{E}_j \cap E_i).$$

Therefore,

$$\begin{aligned}
\mu(E) &= \sum_{j=1}^{n} \mu(\tilde{E}_j) \\
&\leq \sum_{j=1}^{n} \sum_{i=1}^{\infty} \mu(\tilde{E}_j \cap E_i) \\
&= \sum_{i=1}^{\infty} \sum_{j=1}^{n} \mu(\tilde{E}_j \cap E_i) \\
&\leq \sum_{i=1}^{\infty} \mu(E_i),
\end{aligned}$$

which proves one of the inequalities.

For the apposite inequality we set (recall $E = \bigcup_{i=1}^{\infty} E_i$) $A_n = \bigcup_{i=1}^{n} E_i$ and $C_n = E \cap A_n^c$ so that $E = A_n \cup C_n$ and $A_n, C_n \in \overline{\mathcal{S}}$ and disjoint. Therefore,

$$\begin{aligned}
\overline{\mu}(A) = \overline{\mu}(A_n) + \overline{\mu}(C_n) &= \overline{\mu}(B_1) + \ldots + \overline{\mu}(B_n) + \mu(C_n) \\
&\geq \sum_{i=1}^{n} \mu(B_i),
\end{aligned}$$

with n arbitrary. This proves the other inequality. $\square$

Proof of Lemma 2.2. Set $E = \bigcup\limits_{i=1}^{n} E_i$, then $E_i = \bigcup\limits_{j=1}^{m} E_{ij},\ S_{ij} \in \mathcal{S}$. By assumption (i),

$$\mu(A) = \sum_{ij} \mu(E_{ij}) = \sum_{i=1}^{n} \mu(E_i)$$

proving (a).

For (b), assume $n = 1$. If $E \subset E_1$, then

$$E_1 = E \cup (E_1 \cap E^c), \quad E_1 \cap E^c \in \overline{\mathcal{S}}.$$

$$\overline{\mu}(E) \leq \overline{\mu}(E) + \overline{\mu}(E_1 \cap E^c) = \overline{\mu}(E_1).$$

For $n > 1$, set

$$\bigcup_{i=1}^{n} E_i = \bigcup_{i=1}^{n} (E_i \cap E_1^c \cap \ldots \cap E_{i-1}^c) = \bigcup_{i=1}^{n} F_i.$$

Then

$$E = E \cap \left(\bigcup_{i=1}^{n} E_i \right) = E \cap F_1 \cup \ldots \cup (E \cap F_n).$$

So by (a), $\overline{\mu}(E) = \sum_{i=1}^{n} \overline{\mu}(E \cap F_i)$. Now, the case $n = 1$ gives

$$\begin{aligned}\sum_{i=1}^{n} \overline{\mu}(E \cap F_i) &\leq \sum_{i=1}^{n} \mu(E_i) \\ &\leq \sum_{i=1}^{n} \overline{\mu}(F_i) \\ &= \mu\left(\bigcup_{i=1}^{n} E_i \right),\end{aligned}$$

where the last inequality follows from (a). $\square$

Proof of Theorem 2.1. Let $\mathcal{S} = \{(a, b] \colon -\infty \leq a < b < \infty\}$. Set $F(\infty) = \lim_{x \uparrow \infty} F(x)$ and $F(-\infty) = \lim_{x \downarrow -\infty} F(x)$. These quantities exist since F is increasing. Define for any

$$\mu(a, b] = F(b) - F(a),$$

for any $-\infty \leq a < b \leq \infty$, where $F(\infty) > -\infty$, $F(-\infty) < \infty$. Suppose $(a, b] = \bigcup_{i=1}^{n} (a_i, b_i]$, where the union is disjoint. By relabeling we may assume that

$$a_1 = a$$

$$b_n = b$$

$$a_i = b_{i-1}.$$

Then $\mu(a_i, b_i] = F(b_i) - F(a_i)$ and

$$\begin{aligned}\sum_{i==1}^{n} \mu(a_i, b_i] &= \sum_{i=1}^{n} F(b_i) - F(a_i) \\ &= F(b) - F(a) \\ &= \mu(a, b],\end{aligned}$$

which proves that condition (i) holds.

For (ii), let $-\infty < a < b < \infty$ and $(a, b] \subset \bigcup_{i=1}^{\infty}(a_i, b_i]$ where the union is disjoint. (We can also order them if we want.) By right continuity of F, given $\varepsilon > 0$ there is a $\delta > 0$ such that

$$F(a + \delta) - F(a) < \epsilon,$$

or equivalently,

$$F(a + \delta) < F(a) + \varepsilon.$$

Similarly, there is a $\eta_i > 0$ such that

$$F(b_i + \eta_i) < F(b_i) + \varepsilon 2^{-i},$$

for all i. Now, $\{(a_i,\ b_i + \eta_i)\}$ forms a open cover for $[a + \delta, b]$. By compactness, there is a finite subcover. Thus,

$$[a + \delta, b] \subset \bigcup_{i=1}^{N}(a_i,\ b_i + \eta_i)$$

and

$$(a + \delta, b] \subset \bigcup_{i=1}^{N}(a_i,\ b_i + \eta_i].$$

Therefore by (b) of Lemma 2.2,

$$\begin{aligned}
F(b) - F(a+\delta) &= \mu(a+\delta, b] \\
&\le \sum_{i=1}^{N} \mu(a_i,\ b_i + \eta_i] \\
&= \sum_{i=1}^{N} F(b_i + \eta_i) - F(a_i) \\
&= \sum_{i=1}^{N} \{F(b_i + \eta_i) - F(b_i) + F(b_i) - F(a_i)\} \\
&\le \sum_{i=1}^{N} \varepsilon 2^{-i} + \sum_{i=1}^{\infty} (F(b_i) - F(a_i)) \\
&\le \varepsilon + \sum_{i=1}^{\infty} F(b_i) - F(a_i).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mu(a, b] &= F(b) - F(a) \\
&\le 2\varepsilon + \sum_{i=1}^{\infty} F(b_i) - F(a_i) \\
&= 2\varepsilon + \sum_{i=1}^{\infty} \mu(a_i, b_i],
\end{aligned}$$

proving (ii) provided $-\infty < a < b < \infty$.

If $(a, b] \subset \bigcup_{i=1}^{\infty} (a_i, b_i]$, a and b arbitrary, and $(A, B] \subset (a, b]$ for any $-\infty < A < B < \infty$, we have by above

$$F(B) - F(A) \le \sum_{i=1}^{\infty} (F(b_i) - F(a_i))$$

and the result follows by taking limits. $\square$

If $F(x) = x,\ \mu$ is called the Lebesgue measure on $\mathbb{R}$. If

$$F(x) = \begin{cases} 0, & x \le 0 \\ x, & 0 < x \le 1 \\ 1, & x > 1 \end{cases}$$

the measure we obtain is called the Lebesgue measure on $\Omega = (0,1]$. Notice that $\mu(\Omega) = 1$.

If μ is a probability measure then $F(x) = \mu(-\infty, x]$ and $\lim_{x\to\infty} F(x) = 1$. $\lim_{x\downarrow -\infty} F(x) = 0$.

Problem 2.2. *Let F be the distribution function defined by*

$$F(x) = \begin{cases} 0, & x < -1 \\ 1+x, & -1 \le x < 0 \\ 2+x^2, & 0 \le x < 2 \\ 9, & x \ge 2 \end{cases}$$

and let μ be the Lebesgue–Stieltjes measure corresponding to F. Find $\mu(E)$ for

(i) $E = \{2\}$,

(ii) $E = [-1/2, 3)$,

(iii) $E = (-1, 0] \cup (1, 2)$,

(iv) $E = \{x\colon |x| + 2x^2 > 1\}$.

Proof of Theorem 2.3. For any $E \subset \Omega$ we define $\mu^*(E) = \inf \sum \mu(A_i)$ where the infimum is taken over all sequences of $\{A_i\}$ in $\mathcal{A}$ such that $E \subset \cup A_i$. Let $\mathcal{A}^*$ be the collection of all subsets $E \subset \Omega$ with the property that

$$\mu^*(F) = \mu^*(F \cap E) + \mu^*(F \cap E^c),$$

for all sets $F \subset \Omega$. These two quantities satisfy:

(i) $\mathcal{A}^*$ is a σ–algebra and μ^* is a measure on $\mathcal{A}^*$.

(ii) If $\mu^*(E) = 0$, then $E \in \mathcal{A}^*$.

(iii) $\mathcal{A} \subset \mathcal{A}^*$ and $\mu^*(E) = \mu(E)$, if $E \subset \mathcal{A}$.

We begin the proof of (i)–(iii) with a simple but very useful observation. It follows easily from the definition that $E_1 \subset E_2$ implies $\mu^*(E_1) \le \mu^*(E_2)$ and that

$E \subset \cup_{j=1}^{\infty} E_j$ implies

$$\mu^*(E) \leq \sum_{j=1}^{\infty} \mu^*(E_j).$$

Therefore,

$$\mu^*(F) \leq \mu^*(F \cap E) + \mu^*(F \cap E^c)$$

is always true. Hence, to prove that $E \in \mathcal{A}^*$, we need to verify that

$$\mu^*(F) \geq \mu^*(F \cap E) + \mu^*(F \cap E^c),$$

for all $F \in \Omega$. Clearly by symmetry if $E \in \mathcal{A}^*$ we have $E^c \in \mathcal{A}^*$.

Suppose E_1 and E_2 are in $\mathcal{A}^*$. Then for all $F \subset \Omega$,

$$\begin{aligned}
\mu^*(F) &= \mu^*(F \cap E_1) + \mu^*(F \cap E_1^c) \\
&= (\mu^*(F \cap E_1 \cap E_2) + \mu^*(F \cap E_1 \cap E_2^c)) \\
&+ (\mu^*(F \cap E_1^c \cap E_2) + \mu^*(E \cap E_1^c \cap E_2^c)) \\
&\geq \mu^*(F \cap (E_1 \cup E_2)) + \mu^*(F \cap (E_1 \cup E_2)^c),
\end{aligned}$$

where we used the fact that

$$E_1 \cup E_2 \subset (E_1 \cap E_2) \cup (E_1 \cap E_2^c) \cup (E_1^c \cap E_2)$$

and the subadditivity of μ^* observed above. We conclude that $E_1 \cup E_2 \in \mathcal{A}^*$. That is, $\mathcal{A}^*$ is an algebra.

Now, suppose $E_j \in \mathcal{A}^*$ are disjoint. Let $E = \bigcup_{j=1}^{\infty} E_j$ and $A_n = \bigcup\limits_{j=1}^{n} E_j$. Since $E_n \in \mathcal{A}^*$ we have (applying the definition with the set $F \cap A_n$)

$$\begin{aligned}
\mu^*(F \cap A_n) &= \mu^*(F \cap A_n \cap E_n) + \mu^*(F \cap A_n \cap E_n^c) \\
&= \mu^*(F \cap E_n) + \mu^*(F \cap A_{n-1}) \\
&= \mu^*(F \cap E_n) + \mu(F \cap E_{n-1}) + \mu^*(F \cap A_{n-2}) \\
&= \sum_{j=1}^{n} \mu^*(F \cap E_j).
\end{aligned}$$

Now, the measurability of A_n together with this gives

$$\begin{aligned}\mu^*(F) &= \mu^*(F \cap A_n) + \mu^*(F \cap A_n^c) \\ &= \sum_{j=1}^{n} \mu^*(F \cap E_j) + \mu^*(F \cap A_n^c) \\ &\geq \sum_{j=1}^{n} \mu^*(F \cap E_j) + \mu^*(F \cap E^c).\end{aligned}$$

Let $n \to \infty$ we find that

$$\begin{aligned}\mu^*(F) &\geq \sum_{j=1}^{\infty} \mu^*(F \cap E_j) + \mu^*(F \cap E^c) \\ &\geq \mu^*(\cup_{j=1}^{\infty}(F \cap E_j)) + \mu^*(F \cap E^c) \\ &= \mu^*(F \cap E) + \mu^*(F \cap E^c) \geq \mu^*(F),\end{aligned}$$

which proves that $E \in \mathcal{A}^*$. If we take $F = E$ we obtain

$$\mu^*(E) = \sum_{j=1}^{\infty} \mu^*(E_j).$$

From this we conclude that $\mathcal{A}^*$ is closed under countable disjoint unions and that μ^* is countably additive. Since any countable union can be written as the disjoint countable union, we see that $\mathcal{A}^*$ is a σ algebra and that μ^* is a measure on it. This proves (i).

If $\mu^*(E) = 0$ and $F \subset \Omega$, then

$$\begin{aligned}\mu^*(F \cap E) + \mu^*(F \cap E^c) &= \mu^*(F \cap E^c) \\ &\leq \mu^*(F).\end{aligned}$$

Thus, $E \in \mathcal{A}^*$ and we have proved (ii).

For (iii), let $E \in \mathcal{A}$. Clearly

$$\mu^*(E) \leq \mu(E).$$

Next, if $E \subset \bigcup_{j=1}^{\infty} E_i,\ E_i \in \mathcal{A}$, we have $E = \bigcup_{j=1}^{\infty} \tilde{E}_i$, where $\tilde{E}_j = E \cap \left(E_j \Big\backslash \bigcup_{i=1}^{j-1} E_j\right)$ and these sets are disjoint and their union is E. Since μ is a measure on $\mathcal{A}$, we have

$$\mu(E) = \sum_{j=1}^{\infty} \mu(\tilde{E}_j) \leq \sum_{j=1}^{\infty} \mu(E_j).$$

Since this holds for any countable covering of E by sets in $\mathcal{A}$, we have $\mu(E) \leq \mu^*(E)$. Hence

$$\mu(E) = \mu^*(E), \text{ for all } E \in \mathcal{A}.$$

Next, let $E \in \mathcal{A}$. Let $F \subset \Omega$ and assume $\mu^*(F) < \infty$. For any $\varepsilon > 0$, choose $E_j \in \mathcal{A}$ with $F \subset \bigcup_{j=1}^{\infty} E_j$ and

$$\sum_{j=1}^{\infty} \mu(E_j) \leq \mu^*(F) + \varepsilon.$$

Using again the fact that μ is a measure on $\mathcal{A}$,

$$\begin{aligned}
\mu^*(F) + \varepsilon &\geq \sum_{j=1}^{\infty} \mu(E_j) \\
&= \sum_{j=1}^{\infty} \mu(E_j \cap E) + \sum_{j=1}^{\infty} \mu(E_j \cap E^c) \\
&\geq \mu^*(F \cap E) + \mu^*(F \cap E^c)
\end{aligned}$$

and since $\varepsilon > 0$ is arbitrary, we have that $E \in \mathcal{A}^*$. This completes the proof of (iii).

With (i)–(iii) out of the way, it is clear how to define $\overline{\mu}^*$. Since $A \subset \mathcal{A}^*$, and $\mathcal{A}^*$ is a σ–algebra, $\sigma(\mathcal{A}) \subset \mathcal{A}^*$. Define $\overline{\mu}(E) = \mu^*(E)$ for $E \in \sigma(\mathcal{A})$. This is clearly a measure and it remains to prove that it is unique under the hypothesis of σ–finiteness of μ. First, the construction of the measure μ^* clearly shows that whenever μ is finite or σ–finite, so are the measure μ^* and $\overline{\mu}$.

Suppose there is another measure $\tilde{\mu}$ on $\sigma(\mathcal{A})$ with $\mu(E) = \tilde{\mu}(E)$ for all $E \in \mathcal{A}$. Let $E \in \sigma(\mathcal{A})$ have finite μ^* measure. Since $\sigma(A) \subset \mathcal{A}^*$,

$$\mu^*(E) = \inf \left\{ \sum_{j=1}^{\infty} \mu(E_j) \colon E \subset \bigcup_{j=1}^{\infty} E_j, E_j \in \mathcal{A} \right\}.$$

However, since $\mu(E_j) = \tilde{\mu}(E_j)$, we see that

$$\begin{aligned} \tilde{\mu}(E) &\leq \sum_{j=1}^{\infty} \tilde{\mu}(E_j) \\ &= \sum_{j=1}^{\infty} \mu(E_j). \end{aligned}$$

This shows that

$$\tilde{\mu}(E) \leq \mu^*(E).$$

Now let $E_j \in \mathcal{A}$ be such that $E \subset \cup_{j=1}^{\infty} E_j$ and

$$\sum_{j=1}^{\infty} \mu(E_j) \leq \mu^*(E) + \varepsilon.$$

Set $\tilde{E} = \cup_{j=1}^{\infty} E_j$ and $\tilde{E}_n = \cup_{j=1}^{n} E_j$. Then

$$\begin{aligned} \mu^*(\tilde{E}) &= \lim_{k \to \infty} \\ &= \lim_{k \to \infty} \tilde{\mu}(E_n) \\ &= \tilde{\mu}(\tilde{E}) \end{aligned}$$

Since

$$\mu^*(\tilde{E}) \leq \mu^*(E) + \varepsilon,$$

we have

$$\mu^*(\tilde{E} \backslash E) \leq \varepsilon.$$

Hence,

$$\begin{aligned}\mu^*(E) &\le \mu^*(\tilde{E}) \\ &= \tilde{\mu}(\tilde{E}) \\ &\le \tilde{\mu}(E) + \tilde{\mu}(E\backslash E) \\ &\le \tilde{\mu}(E) + \varepsilon.\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, $\overline{\mu}(E) = \mu^*(E) = \tilde{\mu}^*(E)$ for all $E \in \sigma(\mathcal{A})$ of finite μ^* measure. Since μ^* is σ–finite, we can write any set $E = \cup_{j=1}^{\infty}(\Omega_j \cap E)$ where the union is disjoint and each of these sets have finite μ^* measure. Using the fact that both $\tilde{\mu}$ and $\overline{\mu}$ are measures, the uniqueness follows from what we have done for the finite case. $\square$

What is the difference between $\sigma(\mathcal{A})$ and $\mathcal{A}^*$? To properly answer this question we need the following

Definition 2.5. *The measure space $(\Omega, \mathcal{F}, \mu)$ is said to be complete if whenever $E \in \mathcal{F}$ and $\mu(E) = 0$ then $A \in \mathcal{F}$ for all $A \subset E$.*

By (ii), the measure space $(\Omega, \mathcal{A}^*, \mu^*)$ is complete. Now, if $(\Omega, \mathcal{F}, \mu)$ is a measure space we define $\mathcal{F}^* = \{E \cup N : E \in \mathcal{F}, \text{ and } N \in \mathcal{F}, \mu(N) = 0\}$. We leave the easy exercise to the reader to check that $\mathcal{F}^*$ is a σ–algebra. We extend the measure μ to a measure on $\mathcal{F}^*$ by defining $\mu^*(E \cup N) = \mu(E)$. The measure space $(\Omega, \mathcal{F}^*, \mu^*)$ is clearly complete. This measure space is called the completion of $(\Omega, \mathcal{F}, \mu)$. We can now answer the above question.

Theorem 2.4. *The space $(\Omega, \mathcal{A}^*, \mu^*)$ is the completion of $(\Omega, \sigma(\mathcal{A}), \overline{\mu})$.*

II
INTEGRATION THEORY

§1 Measurable Functions.

In this section we will assume that the space $(\Omega.\mathcal{F},\mu)$ is σ–finite. We will say that the set $A\subset\Omega$ is measurable if $A\in\mathcal{F}$. When we say that $A\subset\mathbb{R}$ is measurable we will always mean with respect to the Borel σ–algebra $\mathcal{B}$ as defined in the last chapter.

Definition 1.1. Let $(\Omega,\mathcal{F})$ be a measurable space. Let f be an extended real valued function defined on Ω. That is, the function f is allowed to take values in $\{+\infty,\infty\}$. f is measurable relative to $\mathcal{F}$ if $\{\omega\in\Omega: f(\omega)>\alpha\}\in\mathcal{F}$ for all $\alpha\in\mathbb{R}$.

Remark 1.1. *When $(\Omega,\mathcal{F},P)$ is a probability space and $f:\Omega\to\mathbb{R}$, we refer to measurable functions as random variables.*

Example 1.1. Let $A\subset\Omega$ be a measurable set. The indicator function of this set is defined by

$$1_A(\omega)=\begin{cases}1 & \text{if } \omega\in A\\ 0 & \text{else.}\end{cases}$$

This function is clearly measurable since

$$\{x\colon 1_A(\omega)<\alpha\}=\begin{cases}\Omega & 1\le\alpha\\ \emptyset & \alpha<0\\ A^c & 0\le\alpha<1.\end{cases}$$

This definition is equivalent to several others as seen by the following

Proposition 1.1. *The following conditions are equivalent.*

(i) $\{\omega : f(\omega) > \alpha\} \in \mathcal{F}$ *for all* $\alpha \in \mathbb{R}$,

(ii) $\{\omega : f(\omega) \leq \alpha\} \in \mathcal{F}$ *for all* $\alpha \in \mathbb{R}$,

(iii) $\{\omega : f(\omega) < \alpha\} \in \mathcal{F}$ *for all* $\alpha \in \mathbb{R}$,

(ii) $\{\omega : f(\omega) \geq \alpha\} \in \mathcal{F}$ *for all* $\alpha \in \mathbb{R}$.

Proof. These follow from the fact that σ–algebras are closed under countable unions, intersections, and complementations together with the following two identities.

$$\{\omega : f(\omega) \geq \alpha\} = \bigcap_{n=1}^{\infty} \{\omega : f(\omega) > \alpha - \frac{1}{n}\}$$

and

$$\{\omega : f(\omega) > \alpha\} = \bigcup_{n=1}^{\infty} \{\omega : f(\omega) \geq \alpha + \frac{1}{n}\} \quad \square$$

Problem 1.1. *Let f be a measurable function on $(\Omega, \mathcal{F})$. Prove that the sets $\{\omega : f(\omega) = +\infty\}$, $\{\omega : f(\omega) = -\infty\}$, $\{\omega : f(\omega) < \infty\}$, $\{\omega : f(\omega) > -\infty\}$, and $\{\omega : -\infty < f(\omega) < \infty\}$ are all measurable.*

Problem 1.2.

(i) *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $f : \Omega \to \mathbb{R}$. Prove that f is measurable if and only if $f^{-1}(E) = \{\omega : f(\omega) \in E\} \in \mathcal{F}$ for every Borel set $E \subset \mathbb{R}$.*

(ii) *With f as in (i) define μ on the Borel sets of $\mathbb{R}$ by $\mu(A) = P\{\omega \in \Omega : f(\omega) \in A\}$. Prove that μ is a probability measure on $(\mathbb{R}, \mathcal{B})$.*

Proposition 1.2. *If f and f_2 are measurable, so are the functions $f_1 + f_2$, $f_1 f_2$, $\max(f_1, f_2)$, $\min(f_1, f_2)$ and cf_1, for any constant c.*

Proof. For the sum note that

$$\{\omega : f_1(\omega) + f_2(\omega)\} = \bigcup \left(\{\omega : f_1(\omega) < r\} \cap \{\omega : f_2(\omega) < \alpha - r\}\right),$$

where the union is taken over all the rational. Again, the fact that countable unions of measurable sets are measurable implies the measurability of the sum. In the same way,

$$\{\omega : \max(f_1(\omega), f_2(\omega)) > \alpha\} = \{\omega : f_1(\omega) > \alpha\} \cup \{\omega : f_2(\omega) > \alpha\}$$

gives the measurability of $\max(f_1, f_2)$. The $\min(f_1, f_2)$ follows from this by taking complements. As for the product, first observe that

$$\{\omega : f_1^2(\omega) > \alpha\} = \{\omega : f_1(\omega) > \sqrt{\alpha}\} \cup \{\omega : f_1(\omega) < -\sqrt{\alpha}\}$$

and hence f_1^2 is measurable. But then writing

$$f_1 f_2 = \frac{1}{2}\left[(f_1 + f_2)^2 - f_1^2 - f_2^2\right],$$

gives the measurability of the product. $\square$

Proposition 1.3. *Let* $\{f_n\}$ *be a sequence of measurable functions on* $(\Omega, \mathcal{F})$, *then* $f = \inf_n f_n,\ \sup_n f_n,\ \limsup f_n,\ \liminf f_n$ *are measurable functions.*

Proof. Clearly $\{\inf_n f_n < \alpha\} = \cup\{f_n < \alpha\}$ and $\{\sup_n f_n > \alpha\} = \cup\{f_n > \alpha\}$ and hence both sets are measurable. Also,

$$\limsup_{n\to\infty} f_n = \inf_n \left\{\sup_{m\geq n} f_m\right\}$$

and

$$\liminf_{n\to\infty} f_n = \sup_n \left(\inf_{m\geq n} f_m\right),$$

the result follows from the first part. $\square$

Problem 1.3. *Let f_n be a sequence of measurable functions. Let $E = \{\omega \in \Omega : \lim f_n(\omega)$ exists$\}$. Prove that E is measurable.*

Problem 1.4. *Let f_n be a sequence of measurable functions converging pointwise to the function f. Proof that f is measurable.*

Proposition 1.4.

(i) Let Ω be a metric space and suppose the collection of all open sets are in the sigma algebra $\mathcal{F}$. Suppose $f : \Omega \to \mathbb{R}$ is continuous. Then f is measurable. In particular, a continuous function $f : \mathbb{R}^n \to \mathbb{R}$ is measurable relative to the Borel σ–algebra in $\mathbb{R}^n$.

(ii) Let $\psi : \mathbb{R} \to \mathbb{R}$ be continuous and $f : \Omega \to \mathbb{R}$ be measurable. Then $\psi(f)$ is measurable.

Proof. These both follows from the fact that for every continuous function f, $\{\omega : f(\omega) > \alpha\} = f^{-1}(\alpha, \infty)$ is open for every α. $\square$

Problem 1.5. *Suppose f is a measurable function. Prove that*

(i) f^p, $p \geq 1$,

(ii) $|f|^p$, $p > 0$,

(iii) $f^+ = \max(f, 0)$,

(iv) $f^- = -\min(f, 0)$

are all measurable functions.

Definition 1.2. Let $f : \Omega \to \mathbb{R}$ be measurable. The sigma algebra generated by f is the sigma algebra in Ω_1 generated by the collection $\{f^{-1}(A) : A \in \mathcal{B}\}$. This is denoted by $\sigma(f)$.

Definition 1.3. A function φ define on $(\Omega, \mathcal{F}, \mu)$ is a simple function if $\varphi(w) = \sum_{i=1}^{n} a_i 1_{A_i}$ where the $A_i's$ are disjoint measurable sets which form a partition of Ω, $(\bigcup A_i = \Omega)$ and the $a_i's$ are constants.

Theorem 1.1. *Let $f\colon \Omega \to [0,\infty]$ be measurable. There exists a sequence of simple functions $\{\varphi_n\}$ on Ω with the property that $0 \le \varphi_1(\omega) \le \varphi_2(\omega) \le \ldots \le f(\omega)$ and $\varphi_n(\omega) \to f(\omega)$, for every $\omega \in \Omega$.*

Proof. Fix $n \ge 1$ and for $i = 1, 2, \ldots, n2^n$, define the measurable sets

$$A_{n_i} = f^{-1}\left[\frac{i-1}{2^n},\ \frac{i}{2^n}\right).$$

Set

$$F_n = f^{-1}([n,\infty])$$

and define the simple functions

$$\varphi_n(\omega) = \sum_{i=1}^{n2^n} \frac{i-1}{2^n} 1_{A_{n_i}}(\omega) + n1_{F_n}(\omega).$$

clearly φ_n is a simple function and it satisfies $\varphi_n(\omega) \le \varphi_{n+1}(\omega)$ and $\varphi_n(\omega) \le f(\omega)$ for all ω.

Fix $\varepsilon > 0$. Let $\omega \in \Omega$. If $f(\omega) < \infty$, then pick n so large that $2^{-n} < \varepsilon$ and $f(\omega) < n$. Then $f(\omega) \in \left[\frac{i-1}{2^n},\ \frac{i}{2^n}\right)$ for some $i = 1, 2, \ldots n2^n$. Thus,

$$\varphi_n(\omega) = \frac{i-1}{2^n}$$

and so,

$$f(x)) - \varphi_n(\omega) < 2^{-n}.$$

By our definition, if $f(\omega) = \infty$ then $\varphi_n(\omega) = n$ for all n and we are done. $\square$

§2 The Integral: Definition and Basic Properties.

Definition 2.1. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space.

(i) If $\varphi(w) = \sum_{i=1}^{n} a_i 1_{A_i}$ is a simple function and $E \in \mathcal{F}$ is measurable, we define the integral of the function φ over the set E by

$$\int\limits_E \varphi(w) d\mu = \sum_{i=1}^{n} a_i \mu(A_i \cap E). \tag{2.1}$$

(We adapt the convention here, and for the rest of these notes, that $0 \cdot \infty = 0$.)

(ii) If $f \geq 0$ is measurable we define the integral of f over the set E by

$$\int_E f d\mu = \sup_{\varphi} \int_E \varphi d\mu, \tag{2.2}$$

where the sup is over all simple functions φ with $0 \leq \varphi \leq f$.

(iii) If f is measurable and at least one of the quantities $\int_E f^+ d\mu$ or $\int_E f^- d\mu$ is finite, we define the integral of f over E to be

$$\int_E f d\mu = \int_E f^+ d\mu - \int_E f^- d\mu.$$

(iv) If

$$\int_E |f| d\mu = \int_E f^+ d\mu + \int_E f^- d\mu < \infty$$

we say that the function f is integrable over the set E. If $E = \Omega$ we denote this collection of functions by $L^1(\mu)$.

We should remark here that since in our definition of simple functions we did not required the constants a_i to be distinct, we may have different representations for the simple functions φ. For example, if A_1 and A_2 are two disjoint measurable sets then $1_{A_1 \cup A_2}$ and $1_{A_1} + 1_{A_2}$ both represents the same simple function. It is clear from our definition of the integral that such representations lead to the same quantity and hence the integral is well defined.

Here are some basic and easy properties of the integral.

Proposition 2.1. *Let f and g be two measurable functions on $(\Omega, \mathcal{F}, \mu)$.*

(i) *If $f \leq g$ on E, then $\int\limits_E f d\mu \leq \int\limits_E g d\mu$.*

(ii) *If $A \subset B$ and $f \geq 0$, then $\int\limits_A f d\mu \leq \int\limits_B f d\mu$.*

(iii) *If c is a constant, then $\int\limits_E c f d\mu = c \int\limits_E f d\mu$.*

(iv) *$f \equiv 0$ on E, then $\int\limits_E f d\mu = 0$ even if $\mu(E) = \infty$.*

(v) *If $\mu(E) = 0$, then $\displaystyle\int\limits_E f d\mu = 0$ even if $f(x) = \infty$ on E.*

(vi) *If $f \geq 0$, then $\displaystyle\int\limits_E f d\mu = \int\limits_\Omega g f_E f d\mu$.*

Proposition 2.2. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Suppose φ and ψ are simple functions.*

(i) *For $E \in \mathcal{F}$ define*

$$\nu(E) = \int\limits_E \varphi d\mu.$$

The ν is a measure on $\mathcal{F}$.

(ii) $\displaystyle\int\limits_\Omega (\varphi + \psi) d\mu = \int\limits_\Omega \varphi d\mu + \int\limits_\Omega \psi d\mu.$

Proof. Let $E_i \in \mathcal{F}$, $E = \bigcup E_i$. Then

$$\begin{aligned}
\nu(E) = \int_E \varphi d\mu &= \sum_{i=1}^n \alpha_i \mu(A_i \cap E) \\
&= \sum_{i=1}^n \alpha_i \sum_{j=1}^\infty \mu(A_i \cap E_j) = \sum_{j=1}^\infty \sum_{i=1}^n \alpha_i \mu(A_i \cap E_j) \\
&= \sum_{i=1}^\infty \nu(E_i).
\end{aligned}$$

By the definition of the integral, $\mu(\emptyset) = 0$. This proves (i). (ii) follows from (i) and we leave it to the reader.

We now come to the "three" important limit theorems of integration: The Lebesgue Monotone Convergence Theorem, Fatou's lemma and the Lebesgue Dominated Convergence theorem.

Theorem 2.1 (Lebesgue Monotone Convergence Theorem). *Suppose* $\{f_n\}$ *is a sequence of measurable functions satisfying:*

(i) $0 \leq f_1(\omega) \leq f_2(\omega) \leq \dots$, *for every* $\omega \in \Omega$,

and

(ii) $f_n(\omega) \uparrow f(\omega)$, *for every* $\omega \in \Omega$.

Then

$$\int\limits_{\Omega} f_n d\mu \uparrow \int f d\mu.$$

Proof. Set

$$\alpha_n = \int\limits_{\Omega} f_n d\mu.$$

Then α_n is nondecreasing and it converges to $\alpha \in [0, \infty]$. Since

$$\int\limits_{\Omega} f_n d\mu \leq \int\limits_{\Omega} f d\mu,$$

for all n we see that if $\alpha = \infty$, then $\int_\Omega f d\mu = \infty$ and we are done. Assume $\int_\Omega f d\mu < \infty$. Since

$$\alpha \leq \int_\Omega f d\mu.$$

we need to prove the opposite inequality. Let $0 \leq \varphi \leq f$ be simple and let $0 < c < 1$. Set

$$E_n = \{\omega : f_n(\omega) \geq c\, \varphi(\omega)\}.$$

Clearly, $E_1 \subset E_2 \subset \dots$ In addition, suppose $\omega \in \Omega$. If $f(\omega) = 0$ then $\varphi(\omega) = 0$ and $\omega \in E_1$. If $f(\omega) > 0$ then $cs(\omega) < f(\omega)$ and since $f_n(\omega) \uparrow f(\omega)$, we have that

$\omega \in E_n$ some n. Hence $\bigcup E_n = \Omega$, or in our notation of Proposition 2.1, Chapter I, $E_n \uparrow \Omega$. Hence

$$\begin{aligned}\int_\Omega f_n d\mu &\geq \int_{E_n} f_n d\mu \\ &\geq c\int_{E_n} \varphi(\omega) d\mu \\ &= c\nu(E_n).\end{aligned}$$

Let $n \uparrow \infty$. By Proposition 2.2 above and Proposition 2.1 of Chapter I,

$$\alpha \geq \lim_{n\to\infty} \int_\Omega f_n d\mu \geq c\nu(\Omega) = c\int_\Omega \varphi d\mu$$

and therefore,

$$\int_{\Omega^E} \varphi d\mu \leq \alpha,$$

for all simple $\varphi \leq f$ and

$$\sup_{\varphi\leq f} \int_\Omega \varphi d\mu \leq \alpha,$$

proving the desired inequality. □

Corollary 2.1. *Let $\{f_n\}$ be a sequence of nonnegative measurable functions and set*

$$f = \sum_{n=1}^{\infty} f_n(\omega).$$

Then

$$\int_\Omega f d\mu = \sum_{n=1}^{\infty} \int_\Omega f_n d\mu.$$

Proof. Apply Theorem 2.1 to the sequence of functions

$$g_n = \sum_{j=1}^{n} f_j.$$

□

Corollary 2.2 (First Borel–Contelli Lemma). *Let $\{A_n\}$ be a sequence of measurable sets. Suppose*

$$\sum_{n=1}^{\infty} \mu(A_n) < \infty.$$

Then $\mu\{A_n, i.o.\} = 0$.

Proof. Let $f(\omega) = \sum_{n=1}^{\infty} 1_{A_n(\omega)}$. Then

$$\begin{aligned}\int_{\Omega} f(\omega) d\mu &= \sum_{n=1}^{\infty} \int_{\Omega} 1_{A_n} d\mu \\ &= \sum_{n=1}^{\infty} \mu(A_n) < \infty.\end{aligned}$$

Thus, $f(\omega) < \infty$ for almost every $\omega \in \Omega$. That is, the set A where $f(\omega) = \infty$ has μ–measure 0. However, $f(\omega) = \infty$ if and only if $\omega \in A_n$ for infinitely many n. This proves the corollary. $\square$

Let μ be the counting measure on $\Omega = \{1, 2, 3, \dots\}$ and define the measurable functions f by $f(j) = a_j$ where a_j is a sequence of nonnegative constants. Then

$$\int_{\Omega} f(j) d\mu(j) = \sum_{j=1}^{\infty} a_j.$$

From this and Theorem 2.1 we have

Corollary 2.3. *Let $a_{ij} \geq 0$ for all i, j. Then*

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{ij} = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} a_{ij}.$$

The above theorem together with theorem 1.1 and proposition 2.2 gives

Corollary 2.4. *Let f be a nonnegative measurable function. Define*

$$\nu(E) = \int_E f d\mu.$$

Then ν is a measure and

$$\int_\Omega g d\nu = \int_\Omega g f d\mu$$

for all nonnegative measurable functions g.

Theorem 2.2 (Fatou's Lemma). *Let $\{f_n\}$ be a sequence of nonnegative measurable functions. Then*

$$\int_\Omega \liminf f_n d\mu \leq \liminf \int_\Omega f_n d\mu.$$

Proof. Set

$$g_n(\omega) = \inf_{m \geq n} f_m(\omega), \ n = 1, 2, \ldots$$

Then $\{g_n\}$ is a sequence of nonnegative measurable functions satisfying the hypothesis of Theorem 2.1. Since

$$\lim_{n\to\infty} g_n(\omega) = \liminf_{n\to\infty} f_n(\omega)$$

and

$$\int g_n d\mu \leq \int f_n d\mu,$$

Theorem 2.1 gives

$$\begin{aligned}
\int_\Omega \liminf_{n\to\infty} f_n d\mu &= \int_\Omega \lim_n g_n d\mu \\
&= \lim_{n\to\infty} \int_\Omega g_n d\mu \\
&\leq \liminf_{n\to\infty} \int_\Omega f_n d\mu.
\end{aligned}$$

This proves the theorem. $\square$

Proposition 2.2. *Let f be a measurable function. Then*

$$|\int_\Omega f d\mu| \leq \int_\Omega |f| d\mu.$$

Proof. We assume the right hand side is finite. Set $\beta = \int\limits_\Omega f d\mu$. Take $\alpha = sign(\beta)$ so that $\alpha\beta = |\beta|$. Then

$$\begin{aligned}\left|\int_\Omega f d\mu\right| &= |\beta| \\ &= \alpha \int_\Omega f d\mu \\ &= \int_\Omega \alpha f d\mu \\ &\leq \int_\Omega |f| d\mu.\end{aligned}$$

Theorem 2.3 (The Lebesgue Dominated Convergence Theorem). *Let $\{f_n\}$ be a sequence of measurable functions such that $f_n(\omega) \to f(\omega)$ for every $\omega \in \Omega$. Suppose there is a $g \in L^1(\mu)$ with $|f_n(\omega)| \leq g(\omega)$. Then $f \in L^1(\mu)$ and*

$$\lim_{n\to\infty} \int_\Omega |f_n - f| d\mu = 0.$$

In particular,

$$\lim_{n\to\infty} \int_\Omega f_n d\mu = \int_\Omega f d\mu.$$

Proof. Since $|f(\omega)| = \lim\limits_{n\to\infty} |f_n(\omega)| \leq g(\omega)$ we see that $f \in L^1(\mu)$. Since $|f_n - f| \leq 2g,\ 0 \leq 2g - |f_n - f|$ and Fatou's Lemma gives

$$\begin{aligned}0 \leq \int_\Omega 2g d\mu &\leq \lim_{n\to\infty} \int_\Omega 2g d\mu + \underline{\lim}\left(-\int_\Omega |f_n - f| d\mu\right) \\ &= \int_\Omega 2g d\mu - \overline{\lim} \int_\Omega |f_n - f| d\mu.\end{aligned}$$

It follows from this that

$$\overline{\lim} \int_{\Omega} |f_n - f| d\mu = 0.$$

Since

$$\left| \int_{\Omega} |f_n d\mu - \int_{\Omega} f d\mu \right| \leq \int_{\Omega} |f_n - f| d\mu,$$

the first part follows. The second part follows from the first. $\square$

Definition 2.2. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Let P be a property which a point $\omega \in \Omega$ may or may not have. We say that P holds almost everywhere on E, and write this as i.e. , if there exists a measurable subset $N \subset E$ such that P holds for all $E \backslash N$ and $\mu(N) = 0$.

For example, we say that $f_n \to f$ almost everywhere if $f_n(\omega) \to f(\omega)$ for all $\omega \in \Omega$ except for a set of measure zero. In the same way, $f = 0$ almost everywhere if $f(\omega) = 0$ except for a set of measure zero.

Proposition 2.3 (Chebyshev's Inequality). *Fix $0 < p < \infty$ and let f be a nonnegative measurable function on $(\Omega, \mathcal{F}, \mu)$. Then for any measurable set e we have*

$$\mu\{\omega \in E \colon \in \colon f(\omega) > \lambda\} \leq \frac{1}{\lambda^p} \int_E f^p d\mu \colon .$$

Proof.

$$\begin{aligned} \lambda^p \mu\{\omega \in E \colon f(\omega) > \lambda\} &= \int_{\{\omega \in E \colon f(\omega) > \lambda\}} \lambda^p d\mu \\ &\leq \int_E f^p d\mu \leq \int_E f^p d\mu \end{aligned}$$

which proves the proposition. $\square$

Proposition 2.4.

(i) Let f be a nonnegative measurable function. Suppose

$$\int_E f d\mu = 0.$$

Then $f = 0$ a.e. on E.

(ii) Suppose $f \in L^1(\mu)$ and $\int_E f d\mu = 0$ for all measurable sets $E \subset \Omega$. Then $f = 0$ a.e. on Ω.

Proof. Observe that

$$\{\omega \in E: f(\omega) > 0\} = \bigcup_{n=1}^{\infty} \{\omega \in \Omega: f(\omega) > 1/n\}.$$

By Proposition 2.3,

$$\mu\{\omega \in E: f(\omega) > 1/n\} \leq n \int_E f d\mu = 0.$$

Therefore, $\mu\{f(\omega) > 0\} = 0$, which proves (i).

For (ii), set $E = \{\omega: f(\omega) \geq 0\} = \{\omega: f(\omega) = f^+(\omega)\}$. Then

$$\int_E f^+ d\mu = \int_E f d\mu = 0,$$

which by (i) implies that $f^+ = 0$, a.e. But then

$$\int_E f d\mu = -\int_E f^- d\mu = 0$$

and this again gives

$$\int_E f^- d\mu = 0$$

which implies that $f^- = 0$, a.e. $\square$

Definition 2.3. The function $\psi\colon (a,b) \to \mathbb{R}$ (the "interval" $(a,b) = \mathbb{R}$ is permitted) is convex if

$$\psi((1-\lambda)x + \lambda y) \leq (1-\lambda)\psi(x) + \lambda\psi(y), \tag{2.3}$$

for all $0 \leq \lambda \leq 1$.

An important property of convex functions is that they are always continuous. This "easy" to see geometrically but the proof is not as trivial. What follows easily from the definition is

Problem 2.1. *Prove that (2.3) is equivalent to the following statement: For all $a < s < t < u < b$,*

$$\frac{\varphi(t) - \varphi(s)}{t-s} \leq \frac{\varphi(u) - \varphi(t)}{u-t}.$$

and conclude that a differentiable function is convex if and only if its derivative is a nondecreasing function.

Proposition 2.5 (Jensen's Inequality). *Let $(\Omega, \mathcal{F}, \mu)$ be a probability space. Let $f \in L^1(\mu)$ and $a < f(\omega) < b$. Suppose ψ is convex on (a,b). The $\psi(f)$ is measurable and*

$$\psi\left(\int_\Omega f d\mu\right) \leq \int_\Omega \psi(f) d\mu.$$

Proof. The measurability of the function $\psi(f)$ follows from the continuity of ψ and the measurability of f using Proposition 1.4. Since $a < f(\omega) < b$ for all $\omega \in \Omega$ and μ is a probability measure, we see that if

$$t = \int\limits_\Omega f d\mu,$$

then $a < t < b$. Let $\ell(x) = c_1 x + c_2$ be the equation of the supporting line of the convex function ψ at the point $(t, \psi(t))$. That is, ℓ satisfies $\ell(t) = \psi(t)$ and

$\psi(x) \geq \ell(x)$ for all $x \in (a,b)$. The existence of such a line follows from Problem 2.1. The for all $\omega \in \Omega$,

$$\psi(f(\omega)) \geq c_1 f(\omega) + c_2 = \ell(f(\omega)).$$

Integrating this inequality and using the fact that $\mu(\Omega) = 1$, we have

$$\begin{aligned}\int_\Omega \psi(f(\omega))d\mu &\geq c_1 \int_\Omega f(\omega)d\mu + c_2 \\ &= \ell\left(\int_\Omega f(\omega)d\mu\right) = \psi\left(\int_\Omega f(\omega)d\mu\right),\end{aligned}$$

which is the desired inequality. $\square$

Examples.

(i) Let $\varphi(x) = e^x$. Then

$$\exp\int_\Omega f d\mu \leq \int_\Omega e^f d\mu.$$

(ii) If $\Omega = \{1, 2, \ldots, n\}$ with the measure μ defined by $\mu\{i\} = 1/n$ and the function f given by $f(i) = x_i$, we obtain

$$\exp\{\frac{1}{n}(x_1 + x_2 + \ldots + x_n)\} \leq \frac{1}{n}\{e^{x_1} + \ldots + e^{x_n}\}.$$

Setting $y_i = e^{x_i}$ we obtain the Geometric mean inequality. That is,

$$(y_1, \ldots, y_n)^{1/n} \leq \frac{1}{n}(y_1 + \ldots + y_n)\ .$$

More generally, extend this example in the following way.

Problem 2.2. *Let $\alpha_1, \cdots, \alpha_n$ be a sequence of positive numbers with $\alpha_1 + \cdots + \alpha_n = 1$ and let $y_1, \cdots, y_n$ be positive numbers. Prove that*

$$y_1^{\alpha_1} \cdots y_n^{\alpha_n} \leq \alpha_1 y_1 + \cdots + \alpha_n y_n.$$

Definition 2.4. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Let $0 < p < \infty$ and set*

$$\|f\|_p = \left(\int_\Omega |f|^p d\mu \right)^{1/p}.$$

We say that $f \in L^p(\mu)$ if $\|f\|_p < \infty$. To define $L^\infty(\mu)$ we set

$$E = \{m \in \mathbb{R}^+ \colon \mu\{\omega \colon |f(\omega)| > m\} = 0\}.$$

If $E = \emptyset$, define $\|f\|_\infty = \infty$. If $E \neq \emptyset$, define $\|f\|_\infty = \inf E$. The function $f \in L^\infty(\mu)$ if $\|f\|_\infty < \infty$.

Suppose $\|f\|_\infty < \infty$. Since

$$f^{-1}(\|f\|_\infty,\ \infty] = \bigcup_{n=1}^{\infty} f^{-1}\left(\|f\|_\infty + \frac{1}{n},\ \infty\right]$$

and $\mu f^{-1}(\|f\|_\infty + \frac{1}{n},\ \infty] =$, we see $\|f\|_\infty \in E$. The quantity $\|f\|_\infty$ is called the essential supremum of f.

Theorem 2.4.

(i) (Hölder's inequality) Let $1 \leq p \leq \infty$ and let q be its conjugate exponent. That is, $\frac{1}{p} + \frac{1}{q} = 1$. If $p = 1$ we take $q = \infty$. Also note that when $p = 2$, $q = 2$. Let $f \in L^p(\mu)$ and $g \in l^q(\mu)$. Then $fg \in L^1(\mu)$ and

$$\int |fg| d\mu \leq \|f\|_p \|g\|_q.$$

(ii) (Minkowski's inequality) Let $1 \leq p \leq \infty$. Then

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

Proof. If $p = 1$ and $q = \infty$, or $q = 1$ and $p = \infty$, we have $|fg(\omega)| \leq \|g\|_\infty |f(\omega)|$. This immediately gives the result when $p = 1$ or $p = \infty$. Assume $1 < p < \infty$ and (without loss of generality) that both f and g are nonnegative. Let

$$A = \left(\int_\Omega f^p d\mu \right)^{1/p} \quad \text{and} \quad B = \left(\int g^q d\mu \right)^{1/q}$$

If $A = 0$, then $f = 0$ almost everywhere and if $B = 0$, then $g = 0$ almost everywhere and in either case the result follows. Assume $0 < A < \infty$ and same for B. Put $F = f/A > 0,\ G = g/B$. Then

$$\int_\Omega F^p d\mu = \int_\Omega G^p d\mu = 1$$

and by Problem 2.2

$$F(\omega) \cdot G(\omega) \le \frac{1}{p} F^p + \frac{1}{q} G^q.$$

Integrating both sides of this inequality gives

$$\int F(\omega) G(\omega) \le \frac{1}{p} + \frac{1}{q} = 1,$$

which implies the (i) after multiplying by $A \cdot B$.

For (ii), the cases $p = 1$ and $p = \infty$ are again clear. Assume therefore that $1 < p < \infty$. As before, we may assume that both f and G are nonnegative. We start by observing that since the function $\psi(x) = x^p$, $x \in \mathbb{R}^+$ is convex, we have

$$\left(\frac{f+g}{2} \right)^p \le \frac{1}{2} f^p + \frac{1}{2} g^p.$$

This gives

$$\int_\Omega (f+g)^p d\mu \le 2^{-(p-1)} \int_\Omega f^p d\mu + 2^{-(p-1)} \int_\Omega g^p d\mu.$$

Thus, $f + g \in L^p(d\mu)$. Next,

$$(f+g)^p = (f+g)(f+g)^{p-1} = f(f+g)^{p-1} + g(f+g)^{p-1}$$

together with Hölder's inequality and the fact that $q(p-1) = p$, gives

$$\begin{aligned} \int_\Omega f(f+g)^{p-1} d\mu &\le \left(\int_\Omega f^p d\mu \right)^{1/p} \left(\int_\Omega (f+g)^{q(p-1)} d\mu \right)^{1/q} \\ &= \left(\int_\Omega f^p d\mu \right)^{1/p} \left(\int_\Omega (f+g)^p d\mu \right)^{1/q}. \end{aligned}$$

In the same way,

$$\int_\Omega g(f+g)^p d\mu \leq \left(\int_\Omega g^p d\mu\right)^{1/p} \left(\int_\Omega (f+g)^p d\mu\right)^{1/q}.$$

Adding these inequalities we obtain

$$\int_\Omega (f+g)^p d\mu \leq \left\{\left(\int_\Omega f^p d\mu\right)^{1/p} + \left(\int_\Omega g^p d\mu\right)^{1/p}\right\}\left\{\int_\Omega (f+g)^p d\mu\right\}^{1/q}.$$

Since $f+g \in L^p(d\mu)$, we may divide by the last expression in brackets to obtain the desired inequality. $\square$

For $f,\ g \in L^p(\mu)$ define $d(f,g) = \|f-g\|_p$. For $1 \leq p \leq \infty$, Minkowski's inequality shows that this function satisfies the triangle inequality. That is,

$$\begin{aligned} d(f,g) = \|f-g\|_p &= \|f-h+h-g\|_p \\ &\leq \|f-h\|_p + \|h-g\|_p \\ &= d(f,h) + d(h,g), \end{aligned}$$

for all $f,g,h \in L^p(\mu)$. It follows that $L^p(\mu)$ is a metric space with respect to $d(\cdot,\cdot)$.

Theorem 2.4. *$L^p(\mu)$, $1 \leq p \leq \infty$, is complete with respect to $d(\cdot\ ,\ \cdot)$.*

Lemma 2.1. *Let g_k be a sequence of functions in L^p and $(0 < p < \infty)$ satisfying*

$$\|g_k - g_{h+1}\|_p \leq \left(\frac{1}{4}\right)^k, \ k = 1, 2, \ldots.$$

Then $\{g_k\}$ converges a.e.

Proof. Set

$$A_k = \{\omega\colon |g_k(\omega) - g_{k+1}(\omega)| > 2^{-k}\}.$$

By Chebyshev's inequality,

$$\begin{aligned} \mu\{A_n\} &\leq 2^{kp}\int_\Omega |g_k - g_{k+1}|^p d\mu \\ &\leq \left(\frac{1}{4}\right)^{kp} \cdot 2^{kp} \\ &= \frac{1}{2^{kp}} \end{aligned}$$

This shows that

$$\sum \mu(A_n) < \infty.$$

By Corollary 2.2, $\mu\{A_n \ i.o.\} = 0$. Thus, for almost all $\omega \in \{A_n \ i.o.\}^c$ there is an $N = N(\omega)$ such that

$$|g_k(\omega) - g_{k+1}(\omega)| \leq 2^{-k},$$

for all $k > N$. It follows from this that $\{g_k(\omega)\}$ is Cauchy in $\mathbb{R}$ and hence $\{ g_k(\omega)\}$ converges. $\square$

Lemma 2.2. *The sequence of functions $\{f_n\}$ converges to f in L^∞ if and only if there is a set measurable set A with $\mu(A) = 0$ such that $f_n \to f$ uniformly on A^c. Also, the sequence $\{f_n\}$ is Cauchy in L^∞ if and only if there is a measurable set A with $\mu(A) = 0$ such that $\{f_n\}$ is uniformly Cauchy in A^c.*

Proof. We proof the first statement, leaving the second to the second to the reader. Suppose $\|f_n - f\|_\infty \to 0$. Then for each $k > 1$ there is an $n > n(k)$ sufficiently large so that $\|f_n - f\|_\infty < \frac{1}{k}$. Thus, there is a set A_k so such that $\mu(A_k) = 0$ and $|f_n(\omega) - f(\omega)| < \frac{1}{k}$ for every $\omega \in A_k^c$. Let $A = \cup A_k$. Then $\mu(A) = 0$ and $f_n \to f$ uniformly on A^c. For the converse, suppose $f_n \to f$ uniformly on A^c and $\mu(A) = 0$. Then given $\varepsilon > 0$ there is an N such that for all $n > N$ and $\omega \in A^c$, $|f_n(\omega) - f(\omega)| < \varepsilon$. This is the same as saying that $\|f_n - f\|_\infty < \varepsilon$ for all $n > N$. $\square$

Proof of Theorem 2.4. Now, suppose $\{f_n\}$ is Cauchy in $L^p(\mu)$. That is, given any $\varepsilon > 0$, there is a N such that for all $n, m \geq N$, $d(f_n, f_m) = \|f_n - f_m\| < \varepsilon$ for all $n, m > N$. Assume $1 \leq p < \infty$. The for each $k = 1, 2, \ldots$, there is a n_k such that

$$\|f_n - f_m\|_p \leq (\frac{1}{4})^k$$

for all $n, m \geq n_k$. Thus, $f_{n_k}(\omega) \to f$ a.e., by Lemma 2.1. We need to show that $f \in L^p$ and that it is the $L^p(\mu)$ limit of $\{f_n\}$. Let $\varepsilon > 0$. Take N so large that

$\|f_n - f_m\|_p < \varepsilon$ for all $n, m \geq N$. Fix such an m. Then by the pointwise convergence of the subsequence and by Fatou's Lemma we have

$$\begin{aligned}\int_\Omega |f - f_m|^p d\mu &= \int_\Omega \lim_{k\to\infty} |f_{n_k} - f_m|^p d\mu \\ &\leq \liminf_{k\to\infty} \int_\Omega |f_{n_k} - f_m|^p \\ &< \varepsilon^p.\end{aligned}$$

Therefore $f_n \to f$ in $L^p(\mu)$ and

$$\int_\Omega |f - f_m|^p d\mu < \infty$$

for m sufficiently large. But then,

$$\|f\|_p = \|f_m - f_m - f\|_p \leq \|f_m\|_p + \|f_m - f\|_p,$$

which shows that $f \in L^p(\mu)$.

Now, suppose $p = \infty$. Let $\{f_n\}$ be Cauchy in L^∞. There is a set A with $\mu(A) = 0$ such that f_n is uniformly Cauchy on A^c, by Lemma 2.2. That is, given $\varepsilon > 0$ there is an N such that for all $n, m > N$ and all $\omega \in A^c$,

$$|f_n(\omega) - f_m(\omega)| < \varepsilon.$$

Therefore the sequence $\{f_n\}$ converges uniformly on A^c to a function f. Define $f(\omega) = 0$ for $\omega \in A$. Then f_n converges to f in $L^\infty(\mu)$ and $f \in L^\infty(\mu)$. $\square$

In the course of proving Theorem 2.4 we proved that if a sequence of functions in L^p, $1 \leq p < \infty$ converges in $L^p(\mu)$, then there is a subsequence which converges a.e. This result is of sufficient importance that we list it here as a corollary.

Corollary 2.5. *Let $f_n \in L^p(\mu)$ with $1 \leq p < \infty$ and $f_n \to f$ in $L^p(\mu)$. Then there exists a subsequence $\{f_{n_k}\}$ with $f_{n_k} \to f$ a.e. as $k \to \infty$.*

The following Proposition will be useful later.

Proposition 2.6. *Let $f \in L^1(\mu)$. Given $\varepsilon > 0$ there exists a $\delta > 0$ such that*

$$\int_E |f| d\mu < \varepsilon \text{ whenever } \mu(E) < \delta.$$

Proof. Suppose the statement is false. Then we can find an $\varepsilon > 0$ and a sequence of measurable sets $\{E_n\}$ with

$$\int_{E_n} |f| d\mu \geq \varepsilon$$

and

$$\mu(E_n) < \frac{1}{2^n}$$

Let $A_n = \cup_{j=n}^{\infty} E_j$ and $A = \cap_{n=1}^{\infty} A_n = \{E_n \quad i.o.\}$. Then $\sum \mu(E_n) < \infty$ and by the Borel–Cantelli Lemma, $\mu(A) = 0$. Also, $A_{n+1} \subset A_n$ for all n and since

$$\nu(E) = \int_E |f| d\mu$$

is a finite measure, we have

$$\begin{aligned} \int_A |f| d\mu &= \lim_{n\to\infty} \int_{A_n} |f| d\mu \\ &\geq \lim_{n\to\infty} \int_{E_n} |f| d\mu \\ &\geq \varepsilon. \end{aligned}$$

This is a contradiction since $\mu(A) = 0$ and therefore the integral of any function over this set must be zero. □

§3 Types of convergence for measurable functions.

Definition 3.1. Let $\{f_n\}$ be a sequence of measurable functions on $(\Omega, \mathcal{F}, \mu)$.

(i) $f_n \to f$ in measure if for all $\varepsilon > 0$,

$$\lim_{n\to\infty} \mu\{\omega \in \Omega \colon |f_n(\omega) - f(\omega)| \geq \varepsilon\} = 0.$$

(ii) $f_n \to f$ almost uniformly if given $\varepsilon > 0$ there is a set $E \in \mathcal{F}$ with $\mu(E) < \varepsilon$ such that $f_n \to f$ uniformly on E^c.

Proposition 3.1. *Let $\{f_n\}$ be measurable and $0 < p < \infty$. Suppose $f_n \to f$ in L^p. Then, $f_n \to f$ in measure.*

Proof. By Chebyshev's inequality

$$\mu\{|f_n - f| \geq \varepsilon\} \leq \frac{1}{\varepsilon^p} \int_\Omega |f_n - f|^p d\mu,$$

and the result follows. $\square$

Example 3.1. Let $\Omega = [0,1]$ with the Lebesgue measure. Let

$$f_n(\omega) = \begin{cases} e^n & 0 \leq \omega \leq \frac{1}{n} \\ 0 & \text{else} \end{cases}$$

Then $f_n \to 0$ in measure but $f_n \not\to 0$ in $L^p(\mu)$ for any $0 < p \leq \infty$. To see this simply observe that

$$\|f_n\|_p = \int_0^1 |f_n(x)|^p dx = \frac{1}{n}\, e^{np} \to \infty$$

and that $\|f_n\|_\infty = e^n \to \infty$, as $n \to \infty$.

Proposition 3.2. *Suppose $f_n \to f$ almost uniformly. Then $f_n \to f$ in measure and almost everywhere.*

Proof. Since $f_n \to f$ almost uniformly, given $\varepsilon > 0$ there is a measurable set E such that $\mu(E) < \varepsilon$ and $f_n \to f$ uniformly on E^c. Let $\eta > 0$ be given. There is a $N = N(\eta)$ such that $|f_n(\omega) - f(\omega)| < \eta$ for all $n \geq N$ and for all $\omega \in E^c$. That is, $\{\mu\colon |f_n(\omega) - f(\omega)| \geq \eta\} \subseteq E$, for all $n \geq N$. Hence, for all $n \geq N$,

$$\mu\{|f_n(\omega) - f(\omega)| \geq \eta\} < \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary we see that for all $\eta > 0$,

$$\lim_{n\to\infty} \mu\{|f_n(\omega) - f(\omega)| \geq \eta\} = 0,$$

proving that $f_n \to f$ in measure.

Next, for each k take $A_k \in \mathcal{F}$ with $\mu(A_k) < \frac{1}{k}$ and $f_n \to f$ uniformly on A_k^c. If $E = \cup_{k=1}^\infty A_k^c$, then $f_n \to f$ on E and $\mu(E^c) = \mu\left(\cap_{k=1}^\infty A_k\right) \leq \mu(A_k) < \frac{1}{k}$ for all k. Thus $\mu(E^c) = 0$ and we have the almost everywhere convergence as well. $\square$

Proposition 3.3. *Suppose $f_n \to f$ in measure. Then there is a subsequence $\{f_{n_k}\}$ which converges almost uniformly to f.*

Proof. Since

$$\mu\{|f_n - f_m| \geq \varepsilon\} \leq \mu\{|f_n - f| \geq \varepsilon/2\} + \mu\{|f_n - f| \geq \varepsilon/2\},$$

we see that $\mu\{f_n - f_m| \geq \varepsilon\} \to 0$ as n and $m \to \infty$. For each k, take n_k such that $n_{k+1} > n_k$ and

$$\mu\{|f_n(\omega) - f_m(\omega)| \geq \frac{1}{2^k}\} \leq \frac{1}{2^k}$$

for all $n,\ m \geq n_k$. Setting $g_k = f_{n_k}$ and $A_k = \{\omega \in \Omega \colon |g_{k+1}(\omega) - g_k(\omega)| \geq \frac{1}{2^k}\}$ we see that

$$\sum_{k=1}^{\infty} \mu(A_k) < \infty.$$

By the Borel–Cantelli Lemma, Corollary 2.2, $\mu\{A_n\ i.o.\} = 0$. However, for every $\omega \notin \{A_n\ i.o.\}$, there is an $N = N(\omega)$ such that

$$|g_{k+1}(\omega) - g_k(\omega)| < \frac{1}{2^k}$$

for all $k \geq N$. This implies that the sequence of real numbers $\{g_k(\omega)\}$ is Cauchy and hence it converges to $g(\omega)$. Thus $g_k \to g$ a.e.

To get the almost uniform convergence, set $E_n = \cup_{k=n}^{\infty} A_k$. Then $\mu(E_n) \leq \sum_{k=n}^{\infty} \mu(A_k)$ and this can be made smaller than ε as soon as n is large enough. If $\omega \notin E_n$, then

$$|g_k(\omega) - g_{k+1}(\omega)| < \frac{1}{2^k}$$

for all $k \in \{n, n+1, n+2, \dots\}$. Thus $g_k \to g$ uniformly on E_n^c.

For the uniqueness, suppose $f_n \to f$ in measure. Then $f_{n_k} \to f$ in measure also. Since we also have $f_{n_k} \to g$ almost uniformly clearly, $f_{n_k} \to g$ in measure and hence $f = g$ a.e. This completes the proof. $\square$

Theorem 3.1 (Egoroff's Theorem). *Suppose $(\Omega, \mathcal{F}, \mu)$ is a finite measure space and that $f_n \to f$ a.e. Then $f_n \to f$ almost uniformly.*

Proof. We use Problem 3.2 below. Let $\varepsilon > 0$ be given. For each k there is a $n(k)$ such that if

$$A_k = \bigcup_{n=n(k)}^{\infty} \{\omega \in \Omega \colon |f_n - f| \geq \frac{1}{k}\},$$

then $\mu(A_k) \leq \varepsilon/2^k$. Thus if

$$A = \bigcup_{k=1}^{\infty} A_k,$$

then $\mu(A) \leq \sum_{k=1}^{\infty} \mu(A_k) < \varepsilon$. Now, if $\delta > 0$ take k so large that $\dfrac{1}{k} < \delta$ and then for any $n > n(k)$ and $\omega \notin A$, $|f_n(\omega) - f(\omega)| < \frac{1}{k} < \delta$. Thus $f_n \to f$ uniformly on A^c. $\square$

Let us recall that if $\{y_n\}$ is a sequence a sequence of real numbers then y_n converges to y if and only if every subsequence $\{y_{n_k}$ has a further subsequence $\{y_{n_{k_j}}\}$ which converges to y. For measurable functions we have the following result.

Proposition 3.3. *The sequence of measurable functions $\{f_n\}$ on $(\Omega, \mathcal{F}, \mu)$ converges to f in measure if and only if every subsequence $\{f_{n_k}\}$ contains a further subsequence converging a.e. to f.*

Proof. Let ε_k be a sequence converging down to 0. Then $\mu\{|f_n - f| > \varepsilon_k\} \to 0$, as $n \to \infty$ for each k. We therefore have a subsequence f_{n_k} satisfying

$$\mu\{|f_{n_k} - f| > \varepsilon_k\} \leq \frac{1}{2^k}.$$

Hence,

$$\sum_{k=1}^{\infty} \mu\{|f_{n_k} - f| > \varepsilon_k\} < \infty$$

and therefore by the first Borel–Cantelli Lemma, $\mu\{|f_{n_k} - f| > \varepsilon_k \text{ i.o.}\} = 0$. Thus $|f_{n_k} - f| < \varepsilon_k$ eventually a.e. Thus $f_{n_k} \to f$ a.e.

For the converse, let $\varepsilon > 0$ and put $y_n = \mu\{|f_n - f| > \varepsilon\}$ and consider the subsequence y_{n_k}. If every f_{n_k} subsequence has a subsequence $f_{n_{k_j}}$ such that $f_{n_{k_j}} \to f$ a.e. Then $\{y_{n_k}\}$ has a subsequence. $y_{n_{k_j}} \to 0$. Therefore $\{y_n\}$ converges to 0 and hence That is $f_n \to 0$ in measure. □

Problem 3.1. *Let $\Omega = [0, \infty)$ with the Lebesgue measure and define $f_n(\omega) = 1_{A_n}(\omega)$ where $A_n = \{\omega \in \Omega \colon n \le \omega \le n + \frac{1}{n}\}$. Prove that $f_n \to 0$ a.e., in measure and in $L^p(\mu)$ but that $f_n \not\to 0$ almost uniformly.*

Problem 3.2. *Let $(\Omega, \mathcal{F}, \mu)$ be a finite measure space. Prove that $f_n \to f$ a.e. if and only if for all $\varepsilon > 0$*

$$\lim_{n\to\infty} \mu\left(\bigcup_{k=n}^{\infty} A_k(\varepsilon)\right) = 0$$

where

$$A_k(\varepsilon) = \{\omega \in \Omega \colon |f_k(\omega) - f(\omega)| \ge \varepsilon\}.$$

Problem 3.3.

(i) Give an example of a sequence of nonnegative measurable functions f_n for which we have strict inequality in Fatou's Lemma.

(ii) Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $\{A_n\}$ be a sequence of measurable sets. Recall that $\liminf_n A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$ and prove that

$$\mu\{\liminf_{n\to\infty} A_n\} \le \liminf_{n\to\infty} \mu\{A_n\}.$$

(converges) Suppose f_n is a sequence of nonnegative measurable functions on $(\Omega, \mathcal{F}, \mu)$ which is pointwise decreasing to f. That is, $f_1(\omega) \ge f_2(\omega) \ge \cdots \ge 0$ and

$f_n(\omega) \to f(\omega)$. Is it true that

$$\lim_{n\to\infty} \int_\Omega f_n d\mu = \int_\Omega f d\mu?$$

Problem 3.4. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and suppose $f \in L^1(P)$. Prove that*

$$\lim_{p\to 0} \|f\|_p = \exp\{\int_\Omega \log |f| dP\}$$

where $\exp\{-\infty\}$ is defined to be zero.

Problem 3.5. *Let $(\Omega, \mathcal{F}, \mu)$ be a finite measure space. Prove that the function $\mu\{|f| > \lambda\}$, for $\lambda > 0$, is right continuous and nonincreasing. Furthermore, if f, f_1, f_2 are nonnegative measurable and λ_1, λ_2 are positive numbers with the property that $f \le \lambda_1 f_1 + \lambda_2 f_2$, then for all $\lambda > 0$,*

$$\mu\{f > (\lambda_1 + \lambda_2)\lambda\} \le \mu\{f_1 > \lambda\} + \mu\{f_2 > \lambda\}.$$

Problem 3.6. *Let $\{f_n\}$ be a nondecreasing sequence of measurable nonnegative functions converging a.e. on Ω to f. Prove that*

$$\lim_{n\to\infty} \mu\{f_n > \lambda\} = \mu\{f > \lambda\}.$$

Problem 3.7. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and suppose $\{f_n\}$ is a sequence of measurable functions satisfying*

$$\sum_{n=1}^\infty \mu\{|f_n| > \lambda_n\} < \infty$$

for some sequence of real numbers λ_n. Prove that

$$\limsup_{n\to\infty} \frac{|f_n|}{\lambda_n} \le 1,$$

a.e.

Problem 3.8. *Let $(\Omega, \mathcal{F}, \mu)$ be a finite measure space. Let $\{f_n\}$ be a sequence of measurable functions on this space.*

(i) Prove that f_n converges to f a.e. if and only if for any $\varepsilon > 0$

$$\lim_{m\to\infty} \mu\{|f_n - f_{n'}| > \varepsilon, \textit{for some } n' > n \geq m\} = 0$$

(ii) Prove that $f_n \to 0$ a.e. if and only if for all $\varepsilon > 0$

$$\mu\{|f_n| > \varepsilon, i.o\} = 0$$

(converges) Suppose the functions are nonnegative. Prove that $f_n \to \infty$ a.e. if and only if for all $M > 0$

$$\mu\{f_n < M, i.o.\} = 0$$

Problem 3.9. *Let $\Omega = [0,1]$ with its Lebesgue measure. Suppose $f \in L^1(\Omega)$. Prove that $x^n f \in L^1(\Omega)$ for every $n = 1, 2, \ldots$ and compute*

$$\lim_{n\to\infty} \int_\Omega x^n f(x)dx.$$

Problem 3.10. *Let $(\Omega, \mathcal{F}, \mu)$ be a finite measure space and f a nonnegative real valued measurable function on Ω. Prove that*

$$\lim_{n\to\infty} \int_\Omega f^n d\mu$$

exists, as a finite number, if and only if $\mu\{f > 1\} = 0$.

Problem 3.11. *Suppose $f \in L^1(\mu)$. Prove that*

$$\lim_{n\to\infty} \int_{\{|f|>n\}} f d\mu = 0$$

Problem 3.12. *Let $\Omega = [0,1]$ and let $(\Omega, \mathcal{F}, \mu)$ be a finite measure space and f a measurable function on this space. Let E be the set of all $x \in \Omega$ such that $f(x)$ is an integer. Prove that the set E is measurable and that*

$$\lim_{n\to\infty} \int_\Omega (\cos(\pi f(x))^{2n} d\mu = \mu(E)$$

Problem 3.13. *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Suppose f and g are positive measurable function such that $fg \geq 1$ a.e. on Ω. Prove that*

$$\int_\Omega fg dP \geq 1$$

Problem 3.14. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and suppose $f \in L^1(P)$. Prove that*

$$\lim_{p\to 0} \|f\|_p = \exp\{\int_\Omega \log|f| dP\}$$

where $\exp\{-\infty\}$ *is defined to be zero.*

Problem 3.15. *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Suppose $f \in L^\infty(P)$ and $\|f\|_\infty > 0$. Prove that*

$$\lim_{n\to\infty} \left(\frac{\int_\Omega |f|^{n+1} dP}{\int_\Omega |f|^n dP} \right) = \|f\|_\infty$$

Problem 3.16. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and f_n be a sequence of measurable functions converging to zero in measure. Let F be a bounded uniformly continuous function on $\mathbb{R}$. Prove that*

$$\lim_{n\to\infty} \int_\Omega F(f_n) dP = F(0)$$

Proclaim 3.17. *Let $(\Omega, \mathcal{F}, P)$ be a probability space.*

(i) Suppose $F : \mathbb{R} \to \mathbb{R}$ is a continuous function and $f_n \to f$ in measure. Prove that $F(f_n) \to F(f)$ in measure.

(i) If $f_n \geq 0$ and $f_n \to f$ in measure. Then

$$\int_\Omega f d\mu \leq \underline{\lim} \int_\Omega f_n d\mu.$$

(ii) Suppose $|f_n| \leq g$ where $g \in L^1(\mu)$ and $f_n \to f$ in measure. Then

$$\int_\Omega f d\mu = \lim \int_\Omega f_n d\mu.$$

Problem 3.18. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $f_1, f_2, \cdots, f_n$ be measurable functions. Suppose $1 < p < \infty$. Prove that*

$$\int_\Omega \Big|\frac{1}{n}\sum_{j=1}^n f_j(x)\Big|^p d\mu(x) \leq \frac{1}{n}\int_\Omega \sum_{j=1}^n |f_j(x)|^p d\mu(x)$$

and

$$\int_\Omega \Big|\frac{1}{n}\sum_{j=1}^n f_j(x)\Big|^p d\mu(x) \leq \left(\frac{1}{n}\sum_{j=1}^n \|f_j\|_p\right)^p$$

Problem 3.19. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let f_n be a sequence of measurable functions satisfying $\|f_n\|_p \leq n^{\frac{1}{p}}$, for $2 < p < \infty$. Prove that the sequence $\{\frac{1}{n}f_n\}$ converges to zero almost everywhere.*

Problem 3.20. *Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and that $f \in L^1(P)$ in nonnegative. Prove that*

$$\sqrt{1 + \|f\|_1^2} \leq \int_\Omega \sqrt{1 + f^2} dP \leq 1 + \|f\|_1$$

Problem 3.21. *Compute, justifying all your steps,*

$$\lim_{n\to\infty} \int_0^n \left(1 - \frac{x}{n}\right)^n e^{x/2} dx.$$

Problem 3.22. *Let probtrip be a probability space. Let f be a measurable function with the property that $\|f\|_2 = 1$ and $\|f\|_1 = \frac{1}{2}$. Prove that*

$$\frac{1}{4}(1-\eta)^2 \leq P\{\omega \in \Omega : |f(\omega)| \geq \frac{\eta}{2}\}.$$

III
PRODUCT MEASURES

Our goal in this chapter is to present the essentials of integration in product space. We begin by defining the product measure. Many of the definitions and properties of product measures are, in some sense, obvious. However, we need to be properly state them and carefully prove them so that they may be freely used in the subsequent Chapters.

§1 Definitions and Preliminaries.

Definition 1.1. If X and Y are any two sets, their Cartesian product $X \times Y$ is the set of all order pairs $\{(x,y)\colon x \in X,\ y \in Y\}$.

If $A \subset X,\ B \subset Y,\ A \times B \subset X \times Y$ is called a rectangle. Suppose $(X, \mathcal{A}),\ (X, \mathcal{B})$ are measurable spaces. A measurable rectangle is a set of the form $A \times B,\ A \in \mathcal{A},\ B \in \mathcal{B}$. A set of the form

$$Q = R_1 \cup \ldots \cup R_n,$$

where the R_i are disjoint measurable rectangles, is called an *elementary* sets. We denote this collection by $\mathcal{E}$.

Exercise 1.1. *Prove that the elementary sets form an algebra. That is, $\mathcal{E}$ is closed under complementation and finite unions.*

We shall denote by $\mathcal{A} \times \mathcal{B}$ the σ–algebra generated by the measurable rectangle which is the same as the σ–algebra generated by the elementary sets.

Theorem 1.1. *Let $E \subset X \times Y$ and define the projections*

$$E_x = \{y \in Y : (x, y) \in E\}, \quad \text{and} \quad E^y = \{x \in X : (x, y) \in E\}.$$

If $E \in \mathcal{A} \times \mathcal{B}$, then $E_x \in \mathcal{B}$ and $E^y \in \mathcal{A}$ for all $x \in X$ and $y \in Y$.

Proof. We shall only prove that if $E \in \mathcal{A} \times \mathcal{B}$ then $E_x \in \mathcal{B}$, the case of E^y being completely the same. For this, let Ω be the collection of all sets $E \in \mathcal{A} \times \mathcal{B}$ for which $E_x \in \mathcal{B}$ for every $x \in X$. We show Ω is a σ–algebra containing all measurable rectangles. To see this, note that if

$$E = A \times B$$

then

$$E_x = \begin{cases} B & \text{if } x \in A \\ \emptyset & \text{if } x \notin A. \end{cases}$$

Thus, $E \subset \Omega$. The collection Ω also has the following properties:

(i) $X \times Y \in \Omega$.

(ii) If $E \in \Omega$ then $E^c \in \Omega$.

This follows from the fact that $(E^c)_x = (E_x)^c$, and that $\mathcal{A}$ is a σ–algebras.

(iii) If $E_i \in \Omega$ then $E = \bigcup_{i=1}^{\infty} E_i \in \Omega$.

For (iii), observe that $E_x = \bigcup_{i=1}^{\infty} (E_i)_x$ where $(E_i)_x \in \mathcal{B}$. Once again, the fact that $\mathcal{A}$ is a σ algebras shows that $E \in \Omega$. (i)–(iii) show that Ω is a σ–algebra and the theorem follows. $\square$

We next show that the projections of measurable functions are measurable. Let $f : X \times Y \to \mathbb{R}$. For fix $x \in X$, define $f_x : Y \to \mathbb{R}$ by $f_x(y) = f(x, y)$ with a similar definition for f^y.

In the case when we have several σ–algebras it will be important to clearly distinguish measurability relative to each one of these sigma algebras. We shall use the notation $f \in \sigma(\mathcal{F})$ to mean that the function f is measurable relative to the σ–algebra $\mathcal{F}$.

Theorem 1.2. *Suppose $f \in \sigma(\mathcal{A} \times \mathcal{B})$. Then*

(i) For each $x \in X$, $f_x \in \sigma(\mathcal{B})$

(ii) For each $y \in X$, $f^y \in \sigma(\mathcal{A})$

Proof. Let V be an open set in $\mathbb{R}$. We need to show that $f_x^{-1}(V) \in \mathcal{B}$. Put

$$Q = f^{-1}(V) = \{(x, y) \colon f(x, y) \in V\}.$$

Since $f \in \sigma(\mathcal{A} \times \mathcal{B})$, $Q \in \mathcal{F} \times \mathcal{G}$. However,

$$Q_x = f_x^{-1}(V) = \{y \colon f_x(y) \in V\},$$

and it follows by Theorem 1.1 that $Q_x \in \mathcal{B}$ and hence $f_x \in \sigma(\mathcal{B})$. The same argument proves (ii). $\square$

Definition 1.2. A monotone class $\mathcal{M}$ is a collection of sets which is closed under increasing unions and decreasing intersections. That is:

(i) If $A_1 \subset A_2 \subset \dots$ and $A_i \in \mathcal{M}$, then $\cup A_i \in \mathcal{M}$

(ii) If $B_1 \supset B_2 \supset \dots$ and $B_i \in \mathcal{M}$, then $\cap B_i \in \mathcal{M}$.

Lemma 1.1 (Monotone Class Theorem)**.** *Let $\mathcal{F}_0$ be an algebra of subsets of X and let $\mathcal{M}$ be a monotone class containing $\mathcal{F}_0$. If $\mathcal{F}$ denotes the σ–algebra generated by $\mathcal{F}_0$ then $\mathcal{F} \subset \mathcal{M}$.*

Proof. Let $\mathcal{M}_0$ be the smallest monotone class containing $\mathcal{F}_0$. That is, $\mathcal{M}_0$ is the intersection of all the monotone classes which contain $\mathcal{F}_0$. It is enough to show that $\mathcal{F} \subset \mathcal{M}_0$. By Exercise 1.1, we only need to prove that $\mathcal{M}_0$ is an algebra. First we prove that $\mathcal{M}_0$ is closed under complementation. For this let $\Omega = \{E : E^c \in \mathcal{M}_0\}$. It follows from the fact that $\mathcal{M}_0$ is a monotone class that Ω is also a monotone class and since $\mathcal{F}_0$ is an algebra, if $E \in \mathcal{F}_0$ then $E \in \Omega$. Thus, $\mathcal{M}_0 \subset \Omega$ and this proves it.

Next, let $\Omega_1 = \{E : E \cup F \in \mathcal{M}_0 \text{ for all } F \in \mathcal{F}_0\}$. Again the fact that $\mathcal{M}_0$ is a monotone class implies that Ω_1 is also a monotone class and since clearly $\mathcal{F}_0 \subset \Omega_1$, we have $\mathcal{M}_0 \subset \Omega_1$. Define $\Omega_2 = \{F : F \cup E \in \mathcal{M}_0 \text{ for all } E \in \mathcal{M}_0\}$. Again Ω_2 is a monotone class. Let $F \in \mathcal{F}_0$. Since $\mathcal{M}_0 \in \Omega_1$, if $E \in \mathcal{M}_0$, then $E \cup F \in \mathcal{M}_0$. Thus $\mathcal{F}_0 \subset \Omega_2$ and hence $\mathcal{M}_0 \subset \Omega_2$. Thus, if $E, F \in \mathcal{M}_0$ then $E \cup F \in \mathcal{M}_0$. This shows that $\mathcal{M}_0$ is an algebra and completes the proof. $\square$

Exercise 1.2. *Prove that an algebra $\mathcal{F}$ is a σ–algebra if and only if it is a monotone class.*

Exercise 1.3. *Let $\mathcal{F}_0$ be an algebra and suppose the two measures μ_1 and μ_2 agree on $\mathcal{F}_0$. Prove that they agree on the σ–algebra $\mathcal{F}$ generated by $\mathcal{F}_0$.*

§2 Fubini's Theorem.

We begin this section with a lemma that will allow us to define the product of two measures.

Lemma 2.1. *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be two σ–finite measure spaces. Suppose $Q \in \mathcal{A} \times \mathcal{B}$. If*

$$\varphi(x) = \nu(Q_x) \quad \textit{and} \quad \psi(y) = \mu(Q^y),$$

then

$$\varphi \in \sigma(\mathcal{A}) \quad \textit{and} \quad \psi \in \sigma(\mathcal{B})$$

and

$$\int_X \varphi(x) d\mu(x) = \int_Y \psi(y) d\nu(y). \tag{2.1}$$

Remark 2.1. With the notation of §1 we can write

$$\nu(Q_x) = \int_Y \chi_Q(x,y)d\nu(y) \tag{2.2}$$

and

$$\mu(Q^y) = \int_X \chi_Q(x,y)d\mu(x). \tag{2.3}$$

Thus (2.1) is equivalent to

$$\int_X \int_Y \chi_Q(x,y)d\nu(y)d\mu(x) = \int_Y \int_X \chi_Q(x,y)d\mu(x)d\nu(y).$$

Remark 2.2. Lemma 2.1 allows us to define a new measure $\mu \times \nu$ on $\mathcal{A} \times \mathcal{B}$ by

$$(\mu \times \nu)(Q) = \int_X \nu(Q_x)d\mu(x) = \int_Y \mu(Q^y)d\nu(y). \tag{2.4}$$

To see that this is indeed a measure let $\{Q_j\}$ be a disjoint sequence of sets in $\mathcal{A} \times \mathcal{B}$. Recalling that $(\cup Q_j)_x = \cup (Q_j)_x$ and using the fact that ν is a measure we have

$$\begin{aligned}
(\mu \times \nu)\left(\bigcup_{j=1}^{\infty} Q_j\right) &= \int_X \nu\left(\left(\bigcup_{j=1}^{\infty} Q_j\right)_x\right) d\mu(x) \\
&= \int_X \nu\left(\left(\bigcup_{j=1}^{\infty} Q_{jx}\right)\right) d\mu(x) \\
&= \int_X \sum_{j=1}^{\infty} \nu(Q_{jx})d\mu(x) \\
&= \sum_{j=1}^{\infty} \int_X \nu(Q_{jx})d\mu(x) \\
&= \sum_{j=1}^{\infty} (\mu \times \nu)(Q_j),
\end{aligned}$$

where the second to the last equality follows from the Monotone Convergence Theorem.

Proof of Lemma 2.1. We assume $\mu(X) < \infty$ and $\nu(Y) < \infty$. Let $\mathcal{M}$ be the collection of all $Q \in \mathcal{A} \times \mathcal{B}$ for which the conclusion of the Lemma is true. We will prove that $\mathcal{M}$ is a monotone class which contains the elementary sets; $\mathcal{E} \subset \mathcal{M}$. By Exercise 1.1 and the Monotone class Theorem, this will show that $\mathcal{M} = \mathcal{F} \times \mathcal{G}$. This will be done in several stages. First we prove that rectangles are in $\mathcal{M}$. That is,

(i) Let $Q = A \times B,\ A \in \mathcal{A},\ B \in \mathcal{B}$. Then $Q \in \mathcal{M}$.

To prove (i) observe that

$$Q_x = \begin{cases} B & \text{if } x \in A \\ \emptyset & \text{if } x \notin A. \end{cases}$$

Thus

$$\begin{aligned} \varphi(x) &= \begin{cases} \nu(B) & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases} \\ &= \chi_A(x)\nu(B) \end{aligned}$$

and clearly $\varphi \in \sigma(\mathcal{A})$. Similarly,

$$\psi(y) = 1_B(y)\mu(A) \in \mathcal{B}.$$

Integrating we obtain that

$$\begin{cases} \displaystyle\int_X \varphi(x)d\mu(x) = \mu(A)\nu(B) \\ \displaystyle\int_Y \varphi(y)d\nu(y) = \mu(A)\nu(B), \end{cases}$$

proving (i).

(ii) Let $Q_1 \subset Q_2 \subset \dots,\ Q_j \in \mathcal{M}$. Then $Q = \bigcup\limits_{j=1}^{\infty} Q_j \in \mathcal{M}$.

To prove this, let

$$\varphi_n(x) = \nu((Q_n)_x) = \nu\left(\left(\bigcup_{j=1}^{n} Q_j\right)_x\right)$$

and

$$\psi_n(y) = \mu(Q_n^y) = \mu\left(\left(\bigcup_{j=1}^{n} Q_j\right)^y\right).$$

Then

$$\varphi_n(x) \uparrow \varphi(x) = \nu(Q_x)$$

and

$$\psi_n(x) \uparrow \psi(x) = \mu(Q_x).$$

Since $\varphi_n \in \sigma(\mathcal{A})$ and $\psi_n \in \sigma(\mathcal{B})$, we have $\varphi \in \sigma(\mathcal{A})$ and $\psi \in \sigma(\mathcal{B})$. Also by assumption

$$\int_X \varphi_n(x) d\mu(x) = \int_Y \psi_n(y) d\nu(y),$$

for all n. By Monotone Convergence Theorem,

$$\int_X \varphi(x) d\mu(x) = \int_Y \varphi(y) d\nu(y)$$

and we have proved (ii).

(iii) Let $Q_1 \supset Q_2 \supset \dots ,\ Q_j \in \mathcal{M}$. Then $Q = \bigcap\limits_{j=1}^{\infty} Q_j \in \mathcal{M}$.

The proof of this is the same as (ii) except this time we use the Dominated Convergence Theorem. That is, this time the sequences $\varphi_n(x) = \nu((Q_n)_x)$, $\psi_n(y) = \mu(Q_n^y)$ are both decreasing to $\varphi(x) = \nu(Q_x)$ and $\psi(y) = \mu(Q^y)$, respectively, and since since both measures are finite, both sequences of functions are uniformly bounded.

(iv) Let $\{Q_i\} \in \mathcal{M}$ with $Q_i \cap Q_j = \emptyset$. Then $\bigcup\limits_{j=1}^{\infty} Q_i \in \mathcal{M}$.

For the proof of this, let $\tilde{Q}_n = \bigcup\limits_{i=1}^{n} Q_i$. Then $\tilde{Q}_n \in \mathcal{M}$, since the sets are disjoint. However, the $\tilde{Q}_n's$ are increasing and it follows from (ii) that their union is in $\mathcal{M}$, proving (iv).

It follows from (i)–(iv) that $\mathcal{M}$ is a monotone class containing the elementary sets $\mathcal{E}$. By the Monotone Class Theorem and Exercise 1.1, $\mathcal{A} \times \mathcal{B} = \sigma(\mathcal{E}) = \mathcal{M}$. This proves the Lemma for finite measure and the following exercise does the rest. $\square$

Exercise 2.1. *Extend the proof of Lemma 2.1 to the case of σ–finite measures.*

Theorem 2.1 (Fubini's Theorem)**.** *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be σ–finite measure spaces. Let $f \in \sigma(\mathcal{A} \times \mathcal{B})$.*

(a) (Tonelli) If f is nonnegative and if

$$\varphi(x) = \int_X f_x(y) d\nu(y), \quad \psi(y) = \int_X f^y(x) d\mu(x), \tag{2.5}$$

then

$$\varphi \in \sigma(\mathcal{A}), \ \psi \in \sigma(\mathcal{B})$$

and

$$\int_X \varphi(x) d\mu(x) = \int_{X \times Y} f(x,y) d(\mu \times \nu) = \int_Y \psi(y) d\nu(y).$$

(b) If f is complex valued such that

$$\varphi^*(x) = \int_Y |f(y)|_x d\nu(y) = \int_X |f(x,y)| d\nu(y) < \infty \tag{2.6}$$

and

$$\int_X \varphi^*(x) d\mu(x) < \infty$$

then

$$f \in L^1(\mu \times \nu).$$

and (2.6) holds. A similarly statement holds for y in place of$x$.

(c) If $f \in L^1(\mu \times \nu)$, then $f_x \in L^1(\nu)$ for a.e. $x \in X$, $f^y \in L^1(\mu)$ for a.e. $y \in Y$, the functions defined in (2.5) are measurable and (2.6) holds.

Proof of (a). If $f = \chi_Q$, $Q \in \mathcal{A} \times \mathcal{B}$, the result follows from Lemma 2.1. By linearity we also have the result for simple functions. Let $0 \leq s_1 \leq \dots$ be nonnegative simple functions such that $s_n(x,y) \uparrow f(x,y)$ for every $(x,y) \in X \times Y$. Let

$$\varphi_n(x) = \int_Y (s_n)_x(y) d\nu(y)$$

and

$$\psi_n(y) = \int_X s_n^y(x) d\mu(x).$$

Then

$$\begin{aligned}\int_X \varphi_n(x) d\mu(x) &= \int_{X\times Y} s_n(x,y) d(\mu \times \lambda) \\ &= \int_Y \psi_n(y) d\mu(y).\end{aligned}$$

Since $s_n(x,y) \uparrow f(x,y)$ for every $(x,y) \in X \times Y$, $\varphi_n(x) \uparrow \varphi(x)$ and $\psi_n(y) \uparrow \psi(y)$. The Monotone Convergence Theorem implies the result. Parts (b) and (c) follow directly from (a) and we leave these as exercises. $\square$

The assumption of σ–finiteness is needed as the following example shows.

Example 2.1. $X = Y = [0,1]$ with $\mu =$ the Lebesgue measure and $\nu =$ the counting measure. Let $f(x,y) = 1$ if $x = y$, $f(x,y) = 0$ if $x \neq y$. That is, the function f is the characteristic function of the diagonal of the square. Then

$$\int_X f(x,y) d\mu(x) = 0, \quad \text{and} \quad \int_Y f(x,y) d\nu(y) = 1.$$

Remark 2.1. Before we can integrate the function f in this example, however, we need to verify that it (and hence its projections) is (are) measurable. This can be seen as follows: Set

$$I_j = \left[\frac{j-1}{n}, \frac{j}{n}\right]$$

and

$$Q_n = (I_1 \times I_1) \cup (I_2 \times I_2) \cup \ldots \cup (I_n \times I_n).$$

Then Q_n is measurable and so is $Q = \cap Q_n$, and hence also f .

Example 2.2. Consider the function

$$f(x,y) = \frac{x^2 - y^2}{(x^2+y^2)^2} \quad \text{on} \quad (0,1) \times (0,1).$$

with the $\mu = \nu =$ Lebesgue measure. Then

$$\int_0^1 \int_0^1 f(x,y)dydx = \pi/2$$

but

$$\int_0^1 \int_0^1 f(x,y)dxdy = -\pi/2.$$

The problem here is that $f \notin L^1\{(0,1)\times(0,1)\}$ since

$$\int_0^1 |f(x,y)|dy \geq 1/2x.$$

Let $m_k =$ Lebesgue measure in $\mathbb{R}^k$ and recall that m_k is complete. That is, if $m_k(E) = 0$ then E is Lebesgue measurable. However, $m_1 \times m_1$ is not complete since $\{x\}\times B$, for any set $B \subset \mathbb{R}$, has $m_1 \times m_1-$ measure zero. Thus $m_2 \neq m_1 \times m_1$. What is needed here is the notion of the completion of a measure. We leave the proof of the first two Theorems as exercises.

Theorem 2.2. *If $(X, \mathcal{F}, \mu)$ is a measure space we let*

$$\mathcal{F}^* = \{E \subset X \colon \exists\ A \text{ and } B \in \mathcal{F},\ A \subset E \subset B \text{ and } \mu(B\backslash A) = 0\}.$$

Then $\mathcal{F}^*$ is a σ–algebra and the function μ^* defined on $\mathcal{F}^*$ by

$$\mu^*(E) = \mu(A)$$

is a measure. The measure space (X, m^*, μ^*) is complete. This new space is called the completion of $(X, \mathcal{F}, \mu)$.

Theorem 2.3. *Let m_n be the Lebesgue measure on $\mathbb{R}^n$, $n = r + s$. Then $m_n = (m_r \times m_j)^*$, the completion of the product Lebesgue measures.*

The next Theorem says that as far as Fubini's theorem is concerned, we need not worry about incomplete measure spaces.

Theorem 2.4. *Let $(X, \mathcal{A}, \mu)$, $(Y, \mathcal{B}, \nu)$ be two complete σ–finite measure spaces. Theorems 2.1 remains valid if $\mu\times\nu$ is replaced by $(\mu\times\nu)^*$ except that the functions φ and ψ are defined only almost everywhere relative to the measures μ and ν, respectively.*

Proof. The proof of this theorem follows from the following two facts.

(i) Let $(X, \mathcal{F}, \mu)$ be a measure space. Suppose $f \in \sigma(\mathcal{F}^*)$. Then there is a $g \in \sigma(\mathcal{F})$ such that $f = g$ a.e. with respect to μ.

(ii) Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be two complete and σ–finite measure spaces. Suppose $f \in \sigma((\mathcal{A}\times\mathcal{B})^*)$ is such that $f = 0$ almost everywhere with respect to $\mu\times\nu$. Then for almost every $x \in X$ with respect to μ, $f_x = 0$ a.e. with respect to ν. In particular, $f_x \in \sigma(\mathcal{B})$ for almost every $x \in X$. A similar statement holds with y replacing x.

Let us assume (i) and (ii) for the moment. Then if $f \in \sigma((\mathcal{A}\times\mathcal{B})^*)$ is nonnegative there is a $g \in \sigma(\mathcal{A}\times\mathcal{B})$ such that $f = g$ a.e. with respect to $\mu\times\nu$. Now, apply Theorem 2.1 to g and the rest follows from (ii).

It remains to prove (i) and (ii). For (i), suppose that $f = \chi_E$ where $E \in \mathcal{A}^*$. By definition $A \subset E \subset B$ with $\mu(A\backslash B) = 0$ and A and $B \in \mathcal{A}$. If we set $g = \chi_A$ we have $f = g$ a.e. with respect to μ and we have proved (i) for characteristic function. We now extend this to simple functions and to nonnegative functions in the usual way; details left to the reader. For (ii) let $\Omega = \{(x, y) : f(x, y) \neq 0\}$. Then $\Omega \in (\mathcal{A}\times\mathcal{B})^*$ and $(\mu\times\nu)(\Omega) = 0$. By definition there is an $\tilde{\Omega} \in \mathcal{A}\times\mathcal{B}$ such that $\Omega \subset \tilde{\Omega}$ and $\mu\times\nu(\tilde{\Omega}) = 0$. By Theorem 2.1,

$$\int_X \nu(\tilde{\Omega}_x)d\mu(x) = 0$$

and so $\nu(\tilde{\Omega}_x) = 0$ for almost every x with respect to μ. Since $\Omega_x \subset \tilde{\Omega}_x$ and the space $(Y, \mathcal{B}, \nu)$ is complete, we see that $\Omega_x \in \mathcal{B}$ for almost every $x \in X$ with respect to the measure μ. Thus for almost every $x \in X$ the projection function $f_x \in \mathcal{B}$ and $f_x(y) = 0$ almost everywhere with respect to μ. This completes the proof of (ii) and hence the theorem. $\square$

Exercise 2.3. *Let f be a nonnegative measurable function on $(X, \mathcal{F}, \mu)$. Prove that for any $0 < p < \infty$,*

$$\int_X f(x)^p d\mu(x) = p \int_0^\infty \lambda^{p-1} \mu\{x \in X : f(x) > \lambda\} d\lambda.$$

Exercise 2.4. *Let $(X, \mathcal{F}, \mu)$ be a measure space. Suppose f and g are two nonnegative functions satisfying the following inequality: There exists a constant C such that for all $\varepsilon > 0$ and $\lambda > 0$,*

$$\mu\{x \in X : f(x) > 2\lambda, g(x) \leq \varepsilon\lambda\} \leq C\varepsilon^2 \mu\{x \in X : f(x) > \lambda\}.$$

Prove that

$$\int_X f(x)^p d\mu \leq C_p \int_X g(x)^p d\mu$$

for any $0 < p < \infty$ for which both integrals are finite where C_p is a constant depending on C and p.

Exercise 2.5. *For any $\alpha \in \mathbb{R}$ define*

$$sign(\alpha) = \begin{cases} 1, & \alpha > 0 \\ 0, & \alpha = 0 \\ -1, & \alpha < 0 \end{cases}$$

Prove that

$$0 \leq sign(\alpha) \int_0^y \frac{\sin(\alpha x)}{x} dx \leq \int_0^\pi \frac{\sin(x)}{x} dx \tag{2.7}$$

for all $y > 0$ and that

$$\int_0^\infty \frac{\sin(\alpha x)}{x} dx = \frac{\pi}{2} sign(\alpha) \tag{2.8}$$

and

$$\int_0^\infty \frac{1 - \cos(\alpha x)}{x^2} dx = \frac{\pi}{2} |\alpha|. \tag{2.9}$$

Exercise 2.6. *Prove that*

$$e^{-\alpha} = \frac{2}{\pi}\int_0^\infty \frac{\cos\alpha s}{1+s^2}ds \tag{2.10}$$

for all $\alpha > 0$. Use (2.10), the fact that

$$\frac{1}{1+s^2} = \int_0^\infty e^{-(1+s^2)t}dt,$$

and Fubini's theorem, to prove that

$$e^{-\alpha} = \frac{1}{\sqrt{\pi}}\int_0^\infty \frac{e^{-t}}{\sqrt{t}}e^{-\alpha^2/4t}dt. \tag{2.11}$$

Exercise 2.7. *Let $S^{n-1} = \{x \in \mathbb{R}^n : |x| = 1\}$ and for any Borel set $E \in S^{n-1}$ set $\tilde{E} = \{r\theta : 0 < r < 1, \theta \in A\}$. Define the measure σ on S^{n-1} by $\sigma(A) = n|\tilde{E}|$. Notice that with this definition the surface area ω_{n-1} of the sphere in $\mathbb{R}^n$ satisfies $\omega_{n-1} = n\gamma_n = 2\pi^{\frac{n}{2}}/\Gamma(\frac{n}{2})$ where γ_n is the volume of the unit ball in $\mathbb{R}^n$. Prove (integration in polar coordinates) that for all nonnegative Borel functions f on $\mathbb{R}^n$,*

$$\int_{\mathbb{R}^n} f(x)dx = \int_0^\infty r^{n-1}\left(\int_{S^{n-1}} f(r\theta)d\sigma(\theta)\right)dr.$$

In particular, if f is a radial function, that is, $f(x) = f(|x|)$, then

$$\int_{\mathbb{R}^n} f(x)dx = \frac{2\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})}\int_0^\infty r^{n-1}f(r)dr = n\gamma_n\int_0^\infty r^{n-1}f(r)dr.$$

Exercise 2.8. *Prove that for any $x \in \mathbb{R}^n$ and any $0 < p < \infty$*

$$\int_{S^{n-1}} |\xi\cdot x|^p d\sigma(\xi) = |x|^p\int_{S^{n-1}} |\xi_1|^p d\sigma(\xi)$$

where $\xi\cdot x = \xi_1 x_1 + \cdots + \xi_n x_n$ is the inner product in $\mathbb{R}^n$.

Exercise 2.9. *Let $e_1 = (1, 0, \ldots, 0)$ and for any $\xi \in S^{n-1}$ define θ, $0 \le \theta \le \pi$ such that $e_1 \cdot \xi = \cos\theta$. Prove, by first integrating over $L_\theta = \{\xi \in S^{n-1} : e_1 \cdot \xi = \cos\theta\}$, that for any $1 \le p < \infty$,*

$$\int_{S^{n-1}} |\xi_1|^p d\sigma(\xi) = \omega_{n-1} \int_0^\pi |\cos\theta|^p (\sin\theta)^{n-2} d\theta. \tag{2.12}$$

Use (2.12) and the fact that for any $r > 0$ and $s > 0$,

$$2\int_0^{\frac{\pi}{2}} (\cos\theta)^{2r-1}(\sin\theta)^{2s-1} d\theta = \frac{\Gamma(s)\Gamma(r)}{\Gamma(r+s)}$$

([Ru1, p. 194]) to prove that for any $1 \le p < \infty$

$$\int_{S^{n-1}} |\xi_1|^p d\sigma(\xi) = \frac{2\pi^{\frac{n-1}{2}}\Gamma(\frac{p+1}{2})}{\Gamma(\frac{n+p}{2})} \tag{2.13}$$

IV
RANDOM VARABLES

§1 Some Basics.

From this point on, $(\Omega, \mathcal{F}, \mu)$ will denote a probability space. $X : \Omega \to \mathbb{R}$ is a random variable if X is measurable relative to $\mathcal{F}$. We will use the notation

$$E(X) = \int_\Omega X dP.$$

$E(X)$ is called the expected value of X, or expectation of X. We recall from Problem – that if X is a random variable, then $\mu(A) = \mu_X(A) = P\{X \in A\}$, $A \in \mathcal{B}(\mathbb{R})$, is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. This measure is called the distribution measure of the random variable X. Two random variables X, Y are equally distributed if $\mu_X = \mu_X$. This is often written as $X \overset{d}{=} Y$ or $X \sim Y$.

If we take the set $A = (-\infty, x]$, for any $x \in \mathbb{R}$, the n

$$\mu_X(-\infty, x] = P\{X \le x\} = F_X(x)$$

defines a distribution function, as we saw in Chapter I. We list some additional properties of this distribution function given the fact that $\mu_X(\mathbb{R}) = 1$ and since it arises from the random variable X.

(i) $F_X(b) = F_X(a) = \mu(a, b]$.

(ii) $\lim\limits_{x\to\infty} F_X(x) = 1$, $\lim\limits_{x\to-\infty} F_X(x) = 0$.

(iii) With $F_X(x-) = \lim\limits_{y\uparrow x} F_X(y)$, we see that $F_X(x-) = P(X < x)$.

(iv) $P\{X = x\} = \mu_X\{x\} = F_X(x) - F_X(x-)$.

It follows from (iv) that F is continuous at $x \in \mathbb{R}$ if and only if x is not an atom of the measure μ. That is, if and only if $\mu_X\{x\} = 0$. As we saw in Chapter I, distribution functions are in a one-to-one correspondence with the probability measures in $(\mathbb{R}, \mathcal{B})$. Also, as we have just seen, every random variable gives rise to a distribution function. The following theorem completes this circle.

Theorem 1.1. *Suppose F is a distribution function. Then there is a probability space $(\Omega, \mathcal{F}, P)$ and a random variable X defined on this space such that $F = F_X$.*

Proof. We take $(\Omega, \mathcal{F}, \mu)$ with $\Omega = (0,1)$, $\mathcal{F} =$ Borel sets and P the Lebesgue measure. For each $\omega \in \Omega$, define

$$X(\omega) = \sup\{y : F(y) < \omega\}.$$

We claim this is the desired random variable. Suppose we can show that for each $x \in \mathbb{R}$,

$$\{\omega \in \Omega : X(\omega) \leq x\} = \{\omega \in \Omega : \omega \leq F(x)\}. \tag{1.1}$$

Clearly then X is measurable and also $P\{X(\omega) \leq x\} = F(x)$, proving that $F = F_X$. To prove (1.1) let $\omega_0 \in \{\omega \in \Omega : \omega \leq F(x)\}$. That is, $\omega_0 \leq F(x)$. Then $x \notin \{y : F(y) < \omega_0\}$ and therefore $X(\omega_0) \leq x$. Thus $\{\omega \in \Omega : \omega \leq F(x)\} \subset \{\omega \in \Omega : X(\omega) \leq x\}$.

On the other hand, suppose $\omega_0 > F(x)$. Since F is right continuous, there exists $\epsilon > 0$ such that $F(x+\epsilon) < \omega_0$. Hence $X(\omega) \geq x + \epsilon > x$. This shows that $\omega_0 \notin \{\omega \in \Omega : X\{\omega\} \leq x\}$ and concludes the proof. $\square$

Theorem 1.2. *Suppose X is a random variable and let $G : \mathbb{R} \to \mathbb{R}$ be Borel measurable. Suppose in addition that G is nonnegative or that $E|G(X)| < \infty$. Then*

$$\int_\Omega G(X(\omega))d(\omega) = E(G(X)) = \int_{\mathbb{R}} G(y)d\mu_X(y). \tag{1.2}$$

Proof. Let $B \subset \mathcal{B}(\mathbb{R})$. Then

$$\begin{aligned} E(1_B(X(\omega)) &= P_*\{X \in B\} \\ &= \mu_X(B) = \int_B d\mu_X \\ &= \int_{\mathbb{R}} 1_B(y) d\mu_X(y). \end{aligned}$$

Thus the result holds for indicator functions. By linearity, it holds for simple functions. Now , suppose G is nonnegative. Let φ_n be a sequence of nonnegative simple functions converging pointwise up to G. By the Monotone Convergence Theorem,

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

If $E|G(X)| < \infty$ write

$$G(X(\omega)) = G^+(X(\omega)) - G^-(X(\omega)).$$

Apply the result for nonnegative G to G^+ and G^- and subtract the two using the fact that $E(G(X)) < \infty$. $\square$

More generally let $X_1, X_2, \ldots, X_n$ be n-random variables and define their join distribution by

$$\mu^n(A) = P\{(X_1, X_2, \ldots, X_n) \in A\}, \quad A \in \mathcal{B}(\mathbb{R}^n).$$

μ^n is then a Borel probability measure on $(\mathbb{R}^n, \mathcal{B})$. As before, if $G : \mathbb{R}^n \to \mathbb{R}$ is Borel measurable nonnegative or $E(G(X_1, X_2, \ldots, X_n)) < \infty$, then

$$E(G(X_1(\omega), X_2(\omega), \ldots, X_n(\omega))) = \int_{\mathbb{R}^n} G(x_1, x_2, \ldots, x_n) d\mu^n(x_1, \ldots, x_n).$$

The quantity EX^p, for $1 \le p < \infty$ is called the p–th moment of the random variable X. The case and the variance is defined by $var(X) = E|X - m|^2$ Note that by expanding this quantity we can write

$$\begin{aligned} var(X) &= EX^2 - 2(EX)^2 + (EX)^2 \\ &= EX^2 - (EX)^2 \end{aligned}$$

If we take the function $G(x) = x^p$ then we can write the p–th moments in terms of the distribution as

$$EX^p = \int_{\mathbb{R}} x^p d\mu_X$$

and with $G(x) = (x = m)^2$ we can write the variance as

$$\begin{aligned} var(X) &= \int_{\mathbb{R}} (x-m)^2 d\mu \\ &= \int_{\mathbb{R}} x^2 d\mu_X - m^2. \end{aligned}$$

Now, recall that if f is a nonnegative measurable function on $(\Omega, \mathcal{F}, P)$ then

$$\mu(A) = \int_A f dP$$

defines a new measure on $(\Omega, \mathcal{F})$ and

$$\int_\Omega g\, d\mu = \int_\Omega g f dP. \tag{1.3}$$

In particular, suppose f is a nonnegative borel measurable function in $\mathbb{R}$ with

$$\int_{\mathbb{R}} f(x) dx = 1$$

where here and for the rest of these notes we will simply write dx in place of dm when m is the Lebesgue measure. Then

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

is a distribution function. Hence if $\mu(A) = \displaystyle\int_A f\, dt$, $A \in \mathcal{B}(\mathbb{R})$ then μ is a probability measure and since

$$\mu(a, b] = \int_a^b f(t) dt = F(b) - F(a),$$

for all interval $(a, b]$ we see that $\mu \sim F$ (by the construction in Chapter I). Let X be a random variable with this distribution function. Then by (1.3) and Theorem 1.2,

$$E(g(X)) = \int_{\mathbb{R}} g(x) d\mu(x) = \int_{\mathbb{R}} g(x) f(x) dx. \tag{1.4}$$

Distributions arising from such $f's$ are called absolutely continuous distributions. We shall now give several classical examples of such distributions. The function f is called the density of the random variable associated with the distribution.

Example 1.1. *The Uniform distribution on* $(0, 1)$*.*

$$f(x) = \begin{cases} 1 & x \in (0,1) \\ 0 & x \notin (0,1) \end{cases}$$

Then

$$F(x) = \begin{cases} 0 & x \leq 0 \\ x & 0 \leq x \leq 1 \\ 1 & x \geq 1 \end{cases}.$$

If we take a random variable with this distribution we find that the variance $var(X) = \frac{1}{12}$ and that the mean $m = \frac{1}{2}$.

Example 1.2. *The exponential distribution of parameter* λ*. Let* $\lambda > 0$ *and set*

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0 & else \end{cases}$$

If X is a random variable with associated to this density, we write $X \sim exp(\lambda)$.

$$EX^k = \lambda \int_0^\infty x^k e^{-\lambda x} dx = \frac{k!}{\lambda^k}$$

Example 1.3. *The Cauchy distribution of parameter a. Set*

$$f(x) = \frac{1}{\pi} \frac{a}{a^2 + x^2}$$

We leave it to the reader to verify that if the random variable X has this distribution then $E|X| = \infty$.

Example 1.3. *The normal distribution. Set*

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

The corresponding random variable is the normal distribution. We write $X \sim N(0,1)$ By symmetry,

$$E(X) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x e^{-\frac{x^2}{2}} dx = 0.$$

To compute the variance let us recall first that for any $\alpha > 0$,

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt.$$

We note that

$$\begin{aligned} \int_{\mathbb{R}} x^2 e^{-\frac{x^2}{2}} dx &= 2 \int_0^\infty x^2 e^{-\frac{x^2}{2}} dx = 2\sqrt{2} \int_0^\infty u^{\frac{1}{2}} e^{-u} dx \\ &= 2\sqrt{2}\Gamma(\frac{3}{2}) = 2\sqrt{1}\Gamma(\frac{1}{2}+1) = \frac{k\sqrt{2}}{k}\Gamma(\frac{1}{2}) = \sqrt{2\pi} \end{aligned}$$

and hence $var(X) = 1$. If we take $\sigma > 0$ and $\mu \in \mathbb{R}$, and set

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

we get the normal distribution with mean μ and variance σ and write $X \sim N(\mu, \sigma)$. For this we have $Ex = \mu$ and $\text{var}(X) = \sigma^2$.

Example 1.4. *The gamma distribution arises from*

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)}\lambda^\alpha x^{\alpha-1}e^{-\lambda x} & x \geq 0 \\ 0, \quad x < 0. \end{cases}$$

We write $X \sim \Gamma(\alpha, \lambda)$ *when the random variable* X *has this density.*

Random variables which take only discrete values are appropriately called "discrete random variables." Here are some examples.

Example 1.5. X *is a Bernoulli random variable with parameter* p, $0 < p < 1$, *if* X *takes only two values one with probability* p *and the other with probability* $1-p$. $P(X=1) = p$ *and* $P(X=0) = 1-p$

For this random variable we have

$$EX = p \cdot 1 + (1-p) \cdot 0 = p,$$

$$EX^2 = 1^2 \cdot p = p$$

and

$$var(X) = p - p^2 = p(1-p).$$

Example 1.6. *We say* X *has Poisson distribution of parameter* λ *if*

$$P\{X = k\} = e^{-\lambda}\frac{\lambda^k}{k!} \qquad k = 0, 1, 2, \ldots.$$

For this random variable,

$$EX = \sum_{k=0}^{\infty} k \, \frac{e^{-\lambda}\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda$$

and

$$\mathrm{Var}(X) = EX^2 - \lambda^2 = \sum_{k=0}^{\infty} \frac{k^2 e^{-\lambda}\lambda^k}{k!} - \lambda^2 = \lambda.$$

Example 1.7. *The geometric distribution of parameter p. For $0 < p < 1$ define*

$$P\{X = k\} = p(1-p)^{k-1}, \ \textit{for} \ \ k = 1, 2, \dots$$

The random variable represents the number of independent trial needed to observe an event with probability p. By the geometric series,

$$\sum_{k=1}^{\infty} (1-p)^k = \frac{1}{p}$$

and we leave it to the reader to verify that

$$EN = \frac{1}{p}$$

and

$$var(N) = \frac{1-p}{p^2}.$$

§2 Independence.

Definition 2.1.

(i) The collection $\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_n$ of σ–algebras is said to be independent if whenever $A_1 \in \mathcal{F}_j, A_2 \in \mathcal{F}_2, \dots, A_n \in \mathcal{F}_n$, then

$$P\left(\cap_{j=1}^{n} A_j\right) = \prod_{j=1}^{n} P(A_j).$$

(ii) A collection $\{X_j : 1 \leq j \leq n\}$ of random variables is said to be (totally) independent if for any $\{B_j : 1 \leq j \leq n\}$ of Borel sets in $\mathbb{R}$,

$$P\{X_1 \in B_1, X_1 \in B_2, \dots X_n \in B_n\} = P\{\underset{j=1}{\overset{n}{\cap}} (X_j \in B_j)\} = \prod_{j=1}^{n} P\{X_j \in B_j\}.$$

(iii) The collection of measurable subsets $A_1, A_2, \dots, A_n$ in a σ–algebra $\mathcal{F}$ is independent if for any subset $I \subset \{1, 2, \dots, n\}$ we have

$$P\left\{\bigcap_{j \in I} (A_j)\right\} = \prod_{j \in I} P\{A_j\}$$

Whenever we have a sequence $\{X_1, \ldots, X_n\}$ of independent random variables with the same distribution, we say that the random variables are identically distributed and write this as i.i.d. We note that (iii) is equivalent to asking that the random variables $1_{A_1}, 1_{A_2}, \ldots, 1_{A_3}$ are independent. Indeed, for one direction we take $B_j = \{1\}$ for $j \in I$ and $B_j = \mathbb{R}$ for $j \notin I$. For the other direction the reader is asked to do

Problem 2.1. *Let $A_1, A_2, \ldots, A_n$ be independent. Proof that $A_1^c, A_2^c, \ldots A_n^c$ and $1_{A_1}, 1_{A_2}, \ldots, 1_{A_n}$ are independent.*

Problem 2.2. *Let X and Y be two random variable and set $\mathcal{F}_1 = \sigma(X)$ and $\mathcal{F}_2 = \sigma(Y)$. (Recall that the sigma algebra generated by the random X, denoted $\sigma(X)$, is the sigma algebra generated by the sets $X^{-1}\{B\}$ where B ranges over all Borel sets in $\mathbb{R}$.) Prove that X, Y are independent if and only if $\mathcal{F}_1, \mathcal{F}_2$ are independent.*

Suppose $\{X_1, X_2, \ldots, X_n\}$ are independent and set

$$\mu^n(B) = P\{X_1, \ldots, X_n) \in B\} \qquad B \in \mathcal{B}(\mathbb{R}^n),$$

as in §1. Then with $B = B_1 \times \cdots \times B_n$ we see that

$$\mu^n(B_1 \times \cdots \times B_n) = \prod_{i=1}^{n} \mu_j(B_j)$$

and hence

$$\boxed{\mu^n = \mu_1 \times \cdots \times \mu_n}$$

where the right hand side is the product measure constructed from $\mu_1, \ldots, \mu_n$ as in Chapter III. As we did earlier. Thus for this probability measure on $(\mathbb{R}^n, \mathcal{B})$, the corresponding n-dimensional distribution is

$$F(x) = \prod_{j=1}^{n} F_{X_j}(x_j),$$

where $x = (x_1, x_2, \ldots, x_n))$.

Definition 2.2. Suppose $\mathcal{A} \subset \mathcal{F}$. $\mathcal{A}$ is a π-system if it is closed under intersections: $\mathcal{A}, B \in \mathcal{A} \Rightarrow A \cap B \in \mathcal{A}$. The subcollection $\mathcal{L} \subset \mathcal{F}$ is a λ-system if (i) $\Omega \in \mathcal{L}$, (ii) $A, B \in \mathcal{L}$ and $A \subset B \Rightarrow B \backslash A \in \mathcal{L}$ and (iii) $A_n \in \mathcal{L}$ and $A_n \uparrow A \Rightarrow A \in \mathcal{L}$.

Theorem 2.1. *Suppose $\mathcal{A}$ is a π-system and $\mathcal{L}$ is a λ-system and $A \subset \mathcal{L}$. Then $\sigma(\mathcal{A}) \subset \mathcal{L}$.*

Theorem 2.2. *Let μ and ν be two probability measures on $(\Omega, \mathcal{F})$. Suppose they agree on the π-system $\mathcal{A}$ and that there is a sequence of sets $A_n \in \mathcal{A}$ with $A_n \uparrow \Omega$. Then $\nu = \mu$ on $\sigma(\mathcal{A})$*

Theorem 2.3. *Suppose $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ are independent and π-systems. Then $\sigma(\mathcal{A}_2), \sigma(\mathcal{A}_2), \ldots, \sigma(\mathcal{A}_n)$ are independent.*

Corollary 2.1. *The random variables $X_1, X_2, \ldots, X_n$ are independent if and only if for all $x = \{x_1, \ldots, x_n\}$, $x_i \in (-\infty, \infty]$.*

$$F(x) = \prod_{j=1}^{n} F_{X_j}(x_i), \tag{2.1}$$

where F is the distribution function of the measure μ^n.

Proof. We have already seen that if the random variables are independent then the distribution function F satisfies 2.1. For the other direction let $x \in \mathbb{R}^n$ and set A_i be the sets of the form $\{X_i \leq x_i\}$. Then

$$\{X_i \leq x_i\} \cap \{X_i \leq y_i\} = \{X_i \leq x_i \wedge y_i\} \in \mathcal{A}_i.$$

Therefore the collection $\mathcal{A}_i$ is a π-system. $\sigma(\mathcal{A}_i) = \sigma(X)$. $\square$

Corollary 2.2. $\mu^n = \mu_1 \times \cdots \times \mu_n$.

Corollary 2.3. *Let $X_1, \ldots, X_n$ with $X_i \geq 0$ or $E|X_i| < \infty$ be independent. Then*

$$E\left(\prod_{j=1}^{n} X_i\right) = \prod_{i=1}^{n} E(X_i).$$

Proof. Applying Fubini's Theorem with $f(x_1, \ldots x_n) = x_1 \cdots x_n$ we have

$$\int_{\mathbb{R}^n} (x_1 \cdots x_n) d(\mu_1 \times \cdots \times \mu_n) = \int_{\mathbb{R}} x_1 d\mu_1(a) \cdots \int_{\mathbb{R}^n} x_n d\mu_n(a).$$

□

It follows as in the proof of Corollary 1.3 that if $X_1, \ldots, X_n$ are independent and $g \geq 0$ or if $E|\prod_{j=1}^{n} g(X_i)| < \infty$, then

$$E\left(\prod_{i=1}^{n} g(X_i)\right) = \prod_{i=1}^{n} E(g(X_i)).$$

We warn the reader not to make any inferences in the opposite direction. It may happen that $E(XY) = (E(X)E(Y)$ and yet X and Y may not be independent. Take the two random variables X and Y with joint distributions given by

$X \backslash Y$	1	0	-1
1	0	a	0
0	b	c	b
-1	0	a	0

with $2a + 2b + c = 1$, $a, b, c > 0$. Then $XY = 0$ and $E(X)E(Y) = 0$. Also by symmetry, $EX = EY = 0$. However, the random variables are not independent. Why? Well, observe that $P(X = 1, Y = 1) = 0$ and that $P(X = 1) = P(X = 1, Y = 1) = ab \neq 0$.

Definition 2.2. *If F and G are two distribution functions we define their convolution by*

$$F * G(z) = \int_{\mathbb{R}} F(z - y) d\mu(y)$$

where μ is the probability measure associated with G. The right hand side is often also written as

$$\int_{\mathbb{R}} F(z-y)dG(y).$$

In this notes we will use both notations.

Theorem 2.4. *If X and Y are independent with $X \sim F_X$, $Y \sim G_Y$, then $X+Y \sim F * G$.*

Proof. Let Let us fix $z \in \mathbb{R}$. Define

$$h(x,y) = 1_{(x+y\leq z)}(x,y)$$

Then

$$\begin{aligned}
F_{X+Y}(z) &= P\{X+Y \leq z\} \\
&= E(h(X,Y)) \\
&= \int_{\mathbb{R}^2} h(x,y)d(\mu_X \times \mu_Y)(x,y) \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x,y)d\mu_X(x) \right) d\mu_Y(y) \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} 1_{-\infty,z-y)}(x)\, d\mu_X(x) \right) d\mu_Y(y) \\
&= \int_{\mathbb{R}} \mu_X(-\infty, z-y)d\mu_Y(y) = \int_{\mathbb{R}} F(z-y)dG(y).
\end{aligned}$$

□

Corollary 2.4. *Suppose X has a density f and $Y \sim G$, and X and Y are independent. Then $X+Y$ has density*

$$h(x) = \int_{\mathbb{R}} f(x-y)dG(y).$$

If both X and Y have densities with g denoting the density of Y. Then

$$h(x) = \int f(x-y)g(y)dy.$$

Proof.

$$
\begin{aligned}
F_{X+Y}(z) &= \int_{\mathbb{R}} F(z-y)dG(y) \\
&= \int_{\mathbb{R}} \int_{-\infty}^{z-y} f(x)dxdG(y) \\
&= \int_{\mathbb{R}} \int_{-\infty}^{z} f(u-y)dudG(y) \\
&= \int_{-\infty}^{z} \int_{\mathbb{R}} f(u-y)dG(y)du \\
&= \int_{-\infty}^{z} \left\{ \int_{\mathbb{R}} f(u-y)g(y)dy \right\} du,
\end{aligned}
$$

which completes the proof. □

Problem 2.1. *Let $X \sim \Gamma(\alpha, \lambda)$ and $Y \sim \Gamma(\beta, \lambda)$. Prove that $X+Y \sim \Gamma(\alpha+\beta, \lambda)$.*

§3 Construction of independent random variables.

In the previous section we have given various properties of independent random variables. However, we have not yet discussed their existence. If we are given a finite sequence $F_1, \ldots F_n$ of distribution functions, it is easy to construct independent random variables with this distributions. To do this, let $\Omega = \mathbb{R}^n$ and $\mathcal{F} = \mathcal{B}(\mathbb{R}^n)$. Let P be the measure on this space such that

$$
P((a_1, b_1] \times \cdots \times (a_n, b_n]) = \prod_{j=1}^{n} (F_j(b_j) - F_j(a_j)).
$$

Define the random variables $X_j : \Omega \to \mathbb{R}$ by $X_j(\omega) = \omega_j$, where $\omega = (\omega_1, \ldots, \omega_n)$. Then for any $x_j \in \mathbb{R}$,

$$
P(X_j \leq x_j) = P(\mathbb{R} \times \cdots \times (-\infty, x_j] \times \mathbb{R} \times \cdots \times \mathbb{R}) = F_j(x_j).
$$

Thus $X_j \sim F_j$. Clearly, these random variables are independent by Corollary 2.1. It is, however, extremely important to know that we can do this for infinitely many distributions.

Theorem 3.1. *Let $\{F_j\}$ be a finite <u>or</u> infinite sequence of distribution functions. Then there exists a probability space $(\Omega, \mathcal{F}, P)$ and a sequence of independent random variables X_j on this space with $X_j \sim F_j$.*

Let $N = \{1, 2, \ldots\}$ and let $\mathbb{R}^{\mathbb{N}}$ be the space of infinite sequences of real numbers. That is, $\mathbb{R}^{\mathbb{N}} = \{\omega = (\omega_1, \omega_2, \ldots) : \omega_i \in \mathbb{R}\}$. Let $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$ be the σ-algebra on $\mathbb{R}^{\mathbb{N}}$ generated by the <u>finite dimensional</u> sets. That is, sets of the form $\{\omega \in \mathbb{R}^{\mathbb{N}} : \omega_i \in B_i, 1 \leq i \leq n\}$, $\mathcal{B}_i \in \mathcal{B}(\mathbb{R})$.

Theorem 3.2 (Kolmogovov's Extension Theorem). *Suppose we are given probability measures μ_n on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ which are consistent. That is,*

$$\mu_{n+1}(a_1, b_1] \times \cdots \times (a_n, b_n] \times \mathbb{R}) = \mu_n(a_1, b_1] \times \cdots \times (a_n, b_n].$$

Then there exists a probability measure P on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^N))$ such that

$$P\{\omega : \omega_i \in (a_i, b_i], 1 \leq i \leq n\} = \mu_n((a_1, b_1] \times \cdots \times (a_n, b_n]).$$

The Means above μ_n are consistent. Now, define

$$X_j : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$$

by

$$X_j(\omega) = \omega_j.$$

Then $\{X_j\}$ are independent under the extension measure and $X_j \sim F_j$.

A different way of constructing independent random variables is the following, at least Bernoulli random variables, is as follows. Consider $\Omega = (0, 1]$ and recall that for $x \in (0, 1)$ we can write

$$x = \sum_{n=1}^{\infty} \frac{\epsilon_n}{2^n}$$

where ϵ_n is either 0 or 1. (This representation in actually unique except for x the dyadic nationals.

Problem 3.1. *Define* $X_n(x) = \epsilon_n$. *Prove that the sequence* $\{X_n\}$ *of random variables is independent.*

Problem 3.2. *Let* $\{A_n\}$ *be a sequence of independent sets. Prove that*

$$P\{\bigcap_{j=1}^{\infty} A_j\} = \prod_{j=1}^{\infty} P\{A_j\}$$

and

$$P\{\bigcup_{j=1}^{\infty} A_j\} = 1 - \prod_{j=1}^{\infty}(1 - P\{A_j\})$$

Problem 3.3. *Let* $\{X_1, \ldots, X_n\}$ *be independent random variables with* $X_j \sim F_j$. *Fin the distribution of the random variables* $\max_{1 \le j \le n} X_j$ *and* $\min_{1 \le j \le n} X_j$.

Problem 3.4. *Let* $\{X_n\}$ *be independent random variables and* $\{f_n\}$ *be Borel measurable. Prove that the sequence of random variables* $\{f_n(X_n)\}$ *is independent.*

Problem 3.5. *Suppose* X *and* Y *are independent random variables and that* $X + Y \in L^p(P)$ *for some* $0 < p < \infty$. *Prove that both* X *and* Y *must also be in* $L^p(P)$.

Problem 3.6. *The covariance of two random variables* X *and* Y *is defined by*

$$\begin{aligned} Cov(X, Y) &= E[(X - EX)(Y - EY)] \\ &= E(XY) - E(X)E(Y). \end{aligned}$$

Prove that

$$var(X_1 + X_2 + \cdots + X_n) = \sum_{j=1}^{n} var(X_j) + \sum_{i,j=1, i \neq j}^{n} Cov(X_i, X_j)$$

and conclude that if the random variables are independent then

$$var(X_1 + X_2 + \cdots + X_n) = \sum_{j=1}^{n} var(X_j)$$

V
THE CLASSICAL LIMIT THEOREMS

§1 Bernoulli Trials.

Consider the sequence of independent random variables which arise from tossing a fair coin.

$$X_i = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases}$$

If we use 1 to denote success (=heads) and 0 to denote failure (=tails) and S_n, for the number of successes in n -trials, we can write

$$S_n = \sum_{j=1}^{n} X_j.$$

We can compute and find that the probability of exactly j successes in n trials is

$$\begin{aligned} P\{S_n = j\} &= \binom{n}{j} P\{\text{any specific sequence of } n \text{ trials with exactly } j \text{ heads}\} \\ &= \binom{n}{j} p^j (1-p)^{n-j} \\ &= \binom{n}{j} p^j (1-p)^{n-j} \\ &= \frac{n!}{j!(n-j)!} p^j (1-p)^{n-j}. \end{aligned}$$

This is called Bernoulli's formula. Let us take $p = 1/2$ which represents a fair coin. Then $\frac{S_n}{n}$ denotes the relative frequency of heads in n trials, or, the average number of successes in n trials. We should expect, in the long run, for this to be 1/2. The precise statement of this is

Theorem 1.1 (Bernoulli's "Law of averages," or 'Weak low of large numbers"). *As n–increases, the probability that the average number of successes deviates from $\frac{1}{2}$ by more than any preassigned number tends to zero. That is,*

$$P\{|\frac{S_n}{n} - \frac{1}{2}| > \varepsilon\} \to 0, \text{ as } n \to \infty.$$

Let $x \in [0,1]$ and consider its dyadic representation. That is, write

$$x = \sum_{n=1}^{\infty} \frac{\varepsilon_n}{2^n}$$

with $\varepsilon_n = 0$ or 1. The number x is a normal number if each digit occurs the "right" proportion of times, namely $\frac{1}{2}$.

Theorem 1.2 (Borel 1909). *Except for a set of Lebesgue measure zero, all numbers in $[0,1]$ are normal numbers. That is, $X_n(x) = \varepsilon_n$ and S_n is the partial sum of these random variables, we have $\frac{S_n(x)}{n} \to \frac{1}{2}$ a.s. as $n \to \infty$.*

The rest of this chapter is devoted to proving various generalizations of these results.

§2 L^2 and Weak Laws.

First, to conform to the language of probability we shall say that a sequence of random variables X_n converges almost surely, and write this as a.s., if it converges a.e. as defined in the Chapter II. If the convergence is in measure we shall say that $X_n \to X$ in probability. That is, $X_n \to X$ if for all $\varepsilon > 0$,

$$P\{|X_n - X| > \varepsilon\} \to 0 \text{ as } n \to \infty.$$

We recall that $X_n \to X$ in L^p then $X_n \to X$ in probability and that there is a subsequence $X_{n_k} \to X$ a.s. In addition, recall that by Problem 3.8 in Chapter II, $X_n \to Y$ a.s. iff for any $\varepsilon > 0$,

$$\lim_{m\to\infty} P\{|Y_n - Y| \leq \varepsilon \text{ for all } n \geq m\} = 1 \tag{2.1}$$

or

$$\lim_{m\to\infty} P\{|X_n - X| > \varepsilon \text{ for all } n \geq m\} = 0. \tag{2.2}$$

The proofs of these results are based on a convenient characterization of a.s. convergence. Set

$$\begin{aligned} A_m &= \bigcap_{n=m}^{\infty} \{|X_n - X| \leq \varepsilon\} \\ &= \{|X_n - X| \leq \varepsilon \text{ for all } n \geq m\} \end{aligned}$$

so that

$$A_m^c = \{|X_n - X| > \varepsilon \text{ for some } n \geq m\}.$$

Therefore,

$$\begin{aligned} \{|X_n - X| > \varepsilon \quad i.o.\} &= \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \{|X_n - X| > \varepsilon\} \\ &= \bigcap_{m=1}^{\infty} A_m^c. \end{aligned}$$

However, since $X_n \to X$ a.s. if and only if $|X_n - X| < \varepsilon$, eventually almost surely, we see that $X_n \to X$ a.s. if and only if

$$P\{|X_n - X| > \varepsilon \quad i.o.\} = \lim_{m\to\infty} P\{A_m^c\} = 0. \tag{2.3}$$

Now, (2.1) and (2.2) follow easily from this. Suppose there is a measurable set N with $P(N) = 0$ such that for all $\omega \in \Omega_0 = \Omega \backslash N,\ X_n(\omega_0) \to X(\omega_0)$.

Set

$$A_m(\varepsilon) = \bigcap_{n=m}^{\infty} \{|X_n - X| \leq \varepsilon\} \qquad (2.3.)$$

$A_m(\varepsilon) \subset A_{m+1}(\varepsilon)$. Now, for each $\omega_0 \in \Omega_0$ there exists an $M(\omega_0, \varepsilon)$ such that for all $n \geq M(\omega_0, \varepsilon)$, $|X_n - X| \leq \varepsilon$. Therefore, $\omega \in A_{M(\omega_0,\varepsilon)}$. Thus

$$\Omega_0 \subset \bigcup_{m=1}^{\infty} A_m(\varepsilon)$$

and therefore

$$1 = P(\Omega_0) = \lim_{m\to\infty} P\{A_m(\varepsilon)\},$$

which proves that (2.1) holds. Conversely, suppose (2.1) holds for all $\varepsilon > 0$ and the $A_m(\varepsilon)$ are as in (2.3) and $A(\varepsilon) = \bigcup_{m=1}^{\infty} A_m(\varepsilon)$. Then

$$P\{A(\varepsilon)\} = P\{\bigcup_{m=1}^{\infty} A_m(\varepsilon)\} = 1.$$

Let $\omega_0 \in A(\varepsilon)$. Then for $\omega_0 \in A(\varepsilon)$ there exists $m = m(\omega_0, \varepsilon)$ such that $|X_m - X| \leq \varepsilon$ Let $\varepsilon = \{1/n\}$. Set

$$A = \bigcap_{n=1}^{\infty} A(\frac{1}{n}).$$

Then

$$P(A) = \lim_{n\to\infty} P(A(\frac{1}{n})) = 1$$

and therefore if $\omega_0 \in A$ we have $\omega_0 \in A(1/n)$ for all n. Therefore $|X_n(\omega_0) - X(\omega_0)| < 1/n$ which is the same as $X_n(\omega_0) \to X(\omega_0)$

Theorem 2.1 (L^2–weak law). *Let $\{X_j\}$ be a sequence of uncorrelated random variables. That is, suppose $EX_iX_j = EX_iEX_j$. Assume that $EX_i = \mu$ and that var $(X_i) \leq C$ for all i, where C is a constant. Let $S_n = \sum_{i=1}^{n} X_i$. Then $\frac{S_n}{n} \to \mu$ as $n \to \infty$ in $L^2(P)$ and in probability.*

Corollary 2.1. *Suppose X_i are i.i.d with $EX_i = \mu$, var $(X_i) < \infty$. Then $\dfrac{S_n}{n} \to \mu$ in L^2 and in probability.*

Proof. We begin by recalling from Problem that if X_i are uncorrelated and $E(X_i^2) < \infty$ then $var(X_1 + \ldots + X_n) = var(X_1) + \ldots + var(X_n)$ and that

$var(cX) = c^2 var(X)$ for any constant c. We need to verify that

$$E\left|\frac{S_n}{n} - \mu\right|^2 \to 0.$$

Observe that $E\left(\frac{S_n}{n}\right) = \mu$ and therefore,

$$\begin{aligned}
E\left|\frac{S_n}{n} - \mu\right|^2 &= \text{var}\left(\frac{S_n}{n}\right) \\
&= \frac{1}{n^2}\ \text{var}\ (S_n) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\ \text{var}\ (X_i) \\
&\leq \frac{Cn}{n^2}
\end{aligned}$$

and this last term goes to zero as n goes to infinity. This proves the L^2. Since convergence in L^p implies convergence in probability for any $0 < p < \infty$, the result follows. $\square$

The assumption in Theorem Here is a standard application of the above weak–law.

Theorem 2.2 (The Weierstrass approximation Theorem). *Let f be a continuous function on $[0,1]$. Then there exists a sequence p_n of polynomials such that $p_n \to f$ uniformly on $[0,1]$.*

Proof. Without loss of generality we may assume that $f(0) = f(1) = 0$ for if this is not the case apply the result to $g(x) = f(x) - f(0) - x(f(1) - f(0))$. Put

$$p_n(x) = \sum_{j=0}^{n}\binom{n}{j}x^j(1-x)^{n-j}f(j/n)$$

recalling that

$$\binom{n}{j} = \frac{n!}{j!(n-j)!}.$$

The functions $p_n(x)$ are clearly polynomials. These are called the Bernstein polynomials of degree n associated with f.

Let $X_1, X_2, \ldots$ i.i.d according to the distribution: $P(X_i = 1) = x,\ 0 < x < 1$. $P(X_i = 0) = 1 - x$ so that $E(X_i) = x$ and $var(X_i) = x(1-x)$. The if S_n denotes their partial sums we have from the above calculation that

$$P\{S_n = j\} = \binom{n}{j} x^j (1-x)^{n-j}.$$

Thus

$$E(f(S_n/n)) = \sum_{j=0}^{n} \binom{n}{j} x^j (1-x)^{n-j} f(j/n) = p_n(x)$$

as $n \to \infty$. Also, $S_n/n \to x$ in probability. By Chebyshev's's inequality,

$$\begin{aligned} P\left\{\left|\frac{S_n}{n} - x\right| > \delta\right\} &\leq \frac{1}{\delta^2} \text{ var}\left(\frac{S_n}{n}\right) \\ &= \frac{1}{\delta^2}\frac{1}{n^2} \text{ var}(S_n) \\ &= \frac{x(1-x)}{n\delta^2} \\ &\leq \frac{1}{4n\delta^2} \end{aligned}$$

for all $x \in [0,1]$ since $x(1-x) \leq \frac{1}{4}$. Set $M = \|f\|_\infty$ and let $\varepsilon > 0$. There exists a $\delta > 0$ such that $|f(x) - f(y)| < \varepsilon$ when $|x - y| < \delta$. Thus

$$\begin{aligned} |p_n(x) - f(x)| &= \left|Ef\left(\frac{S_n}{n}\right) - f(x)\right| \\ &= \left|E\left(f\left(\frac{S_n}{n}\right) - f(x)\right)\right| \\ &\leq E\left|f\left(\frac{S_n}{n}\right) - f(x)\right| \\ &= \int_{\{|\frac{S_n}{n} - x| < \delta\}} \left|f\left(\frac{S_n}{n}\right) - f(x)\right| dP \\ &+ \int_{\{|\frac{S_n}{n} - x| \geq \delta} \left|f\left(\frac{S_n}{n}\right) - f(x)\right| dP \\ &< \varepsilon + 2MP\left\{\left|\frac{S_n}{n} - x\right| \geq \delta\right\}. \end{aligned}$$

Now, the right hand side can be made smaller than 2ε by taking n large enough and independent of x. This proves the result. $\square$

The assumption that the variances of the random variables are uniformly bounded can be considerably weaken.

Theorem 2.3. *Let X_i be i.i.d. and assume*

$$\lambda P\{|X_i| > \lambda\} \to 0 \tag{1.3}$$

as $\lambda \to \infty$. Let $S_n = \sum_{j=1}^n X_j$ and $\mu_n = E(X_1 1_{(|X_1| \le n)})$. Then

$$\frac{S_n}{n} - \mu_n \to 0$$

in probability.

Remark 2.1. *The condition (1.3) is necessary to have a sequence of numbers a_n such that $\frac{S_n}{n} - a_n \to 0$ in probability. For this, we refer the reader to Feller, Vol II (1971).*

Before proving the theorem we have a

Corollary 2.2. *Let X_i be i.i.d. with $E|X_1| < \infty$. Let $\mu = EX_i$. Then $\frac{S_n}{n} \to \mu$ in probability.*

Proof of Corollary. First, by the Monotone Convergence Theorem and Chebyshev's's inequality, $\lambda P\{|X_i| > \lambda\} = \lambda P\{|X_1| > \lambda\} \to 0$ as $\lambda \to \infty$ and $\mu_n \to E(X) = \mu$. Hence,

$$\begin{aligned} P\left\{\left|\left(\frac{S_n}{n} - \mu\right)\right| > \varepsilon\right\} &= P\left\{\left|\frac{S_n}{n} - \mu + \mu_n - \mu_n\right| > \varepsilon\right\} \\ &\le P\left\{\left|\frac{S_n}{n} - \mu_n\right| > \varepsilon/2\right\} + P\{\mu_n - \mu| > \varepsilon/2\} \end{aligned}$$

and these two terms go to zero as $n \to \infty$. $\square$

Lemma 2.1 (triangular arrays). *Let* $X_{n,k}, 1 \le k \le n,\ n = 1, 2, \ldots$ *be a triangular array of random variables and assume that for each* n, $X_{n,k}, 1 \le k \le n$, *are independent. Let* $b_n > 0$, $b_n \to \infty$ *as* $n \to \infty$ *and define the truncated random variables by* $\overline{X}_{n,k} = X_{n,k} 1_{|X_{n,k}| \le b_n)}$. *Suppose that*

(i) $\displaystyle\sum_{n=1}^{n} P\{|X_{n,k}| > b_n) \to 0,$ *as* $n \to \infty$.

(ii) $\displaystyle\frac{1}{b_n^2} \sum_{k=1}^{n} E\overline{X}_{n,k}^2 \to 0$ *as* $n \to \infty$.

Put $a_n = \displaystyle\sum_{k=1}^{n} E\overline{X}_{n,k}$ *and set* $S_n = X_{n,1} + X_{n,2} + \ldots + X_{n,n}$. *Then*

$$\frac{S_n - a_n}{b_n} \to 0$$

in probability.

Proof. Let $\overline{S}_n = \overline{X}_{n,1} + \ldots + \overline{X}_{n,n}$. Then

$$\begin{aligned} P\left\{\frac{|S_n - a_n|}{b_n} > \varepsilon\right\} &= P\left\{\left|\frac{S_n - a_n}{b_n}\right| > \varepsilon,\ S_n = \overline{S}_n, S_n \neq \overline{S}_n\right\} \\ &\le P\{S_n \neq \overline{S}_n\} + P\left\{\frac{|\overline{S}_n - a_n|}{b_n} > \varepsilon\right\}. \end{aligned}$$

However,

$$\begin{aligned} P\{S_n \neq \tilde{S}_n\} &\le P\left\{\bigcup_{k=1}^{n} \{\overline{X}_{n,k} \neq X_{n,k}\}\right\} \\ &\le \sum_{k=1}^{n} P\{\overline{X}_{n,k} \neq X_{n,k}\} \\ &= \sum_{k=1}^{n} P\{|X_{n,k}| > b_n\} \end{aligned}$$

and this last term goes to zero by (i).

Since $a_n = E\overline{S}_n$, we have

$$P\left\{\frac{|\overline{S}_n - a_n|}{b_n} > \varepsilon\right\} \le \frac{1}{\varepsilon^2 b_n^2} E|\overline{S}_n - a_n|^2$$
$$= \frac{1}{\varepsilon^2 b_n^2} \sum_{k=1}^{n} var(\overline{X}_{n,k})$$
$$\le \frac{1}{\varepsilon^2 b_n^2} \sum_{k=1}^{n} E\overline{X}_{n,k}^2$$

and this goes to zero by (ii).

Proof of Theorem 2.3. We apply the Lemma with $X_{n,k} = X_k$ and $b_n = n$. We first need to check that this sequence satisfies (i) and (ii). For (i) we have

$$\sum_{k=1}^{n} P\{|X_{n,k}| > n\} = nP\{|X_1| > n\},$$

which goes to zero as $n \to \infty$ by our assumption. For (ii) we see that that since the random variables are i.i.d we have

$$\frac{1}{n^2} \sum_{k=1}^{n} E\overline{X}_{n,1}^2 = \frac{1}{n} E\overline{X}_{n,1}^2.$$

Let us now recall that by Problem 2.3 in Chapter III, for any nonnegative random variable Y and any $0 < p < \infty$,

$$EY^p = p\int_0^{\infty} \lambda^{p-1} P\{Y > \lambda\} d\lambda$$

Thus,

$$E|\overline{X}_{n,1}^2| = 2\int_0^{\infty} \lambda P\{\overline{X}_{n,1}| > \lambda\} d\lambda = 2\int_0^{n} P\{|X_{n,1}| > \lambda\} d\lambda.$$

We claim that as $n \to \infty$,

$$\frac{1}{n}\int_0^{n} \lambda P\{|X_1| > \lambda\} d\lambda \to 0.$$

For this, let

$$g(\lambda) = \lambda P\{|X_1| > \lambda\}.$$

Then $0 \le g(\lambda) \le \lambda$ and $g(\lambda) \to 0$. Set $M = \sup_{\lambda > 0} |g(\lambda)| < \infty$ and let $\varepsilon > 0$. Fix k_0 so large that $g(\lambda) < \varepsilon$ for all $\lambda > k_0$. Then

$$\begin{aligned}\int_0^n \lambda P\{|X_1| > \lambda\} dx &= M + \int_{k_0}^n \lambda P\{|x_1| > \lambda\} d\lambda \\ &< M + \varepsilon(n - k_0).\end{aligned}$$

Therefore

$$\frac{1}{n}\int_0^n \lambda P\{|x_1| > \lambda\} < \frac{M}{n} + \varepsilon\left(\frac{n-k_0}{n}\right).$$

The last quantity goes to ε as $n \to \infty$ and this proves the claim. $\square$

§3 Borel–Cantelli Lemmas.

Before we stay our Borel–Cantelli lemmas for independent events, we recall a few already proven facts. If $A_n \subset \Omega$, then

$$\{A_n,\ i.o.\} = \overline{\lim} A_n = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n$$

and

$$\{A_n, eventually\} = \underline{\lim} A_n = \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} A_n$$

Notice that

$$\overline{\lim} 1_{A_n} = 1_{\{\overline{\lim} A_n\}}$$

and

$$\underline{\lim} 1_{A_n}(\omega) = 1_{\{\underline{\lim} A_n\}}$$

It follows from Fatou's Lemma

$$P(\underline{\lim} A_n) \le \underline{\lim} P\{A_n\}$$

and that

$$\overline{\lim} P\{A_n\} \le P\{\overline{\lim} A_n\}.$$

Also recall Corollary 2.2 of Chapter II.

First Borel–Cantelli Lemma. *If* $\sum\limits_{n=1}^{\infty} p(A_n) < \infty$ *then* $P\{A_n,\ i.o.\} = 0$.

Question: Is it possible to have a converse ? That is, is it true that $P\{A_n,\ i.o.\} = 0$ implies that $\sum_{n=1}^{\infty} P\{A_n\} = \infty$? The answer is no, at least not in general.

Example 3.1. Let $\Omega = (0,1)$ with the Lebesgue measure on its Borel sets. Let $a_n = 1/n$ and set $A_n = (0, 1/n)$. Clearly then $\sum P(A_n) = \infty$. But, $P\{A_n\ i.o.\} = P\{\emptyset\} = 0$.

Theorem 3.1 (The second Borel–Cantelli Lemma). *Let* $\{A_n\}$ *be a sequence of independent sets with the property that* $\sum P(A_n) = \infty$ *Then* $P\{A_n\ i.o.\} = 1$.

Proof. We use the elementary inequality $(1-x) \leq e^{-x}$ valid for $0 \leq x \leq 1$. Let Fix N. By independence,

$$\begin{aligned} P\left\{\bigcap_{n=m}^{N} A_n^c\right\} &= \prod_{n=m}^{N} P\{A_n^c\} \\ &= \prod_{n=m}^{N} \{1 - P\{A_n\}\} \\ &\leq \prod_{n=m}^{N} e^{-P\{A_n\}} \\ &= \exp^{-\{\sum\limits_{n=m}^{N} P\{A_n\}\}} \end{aligned}$$

and this quantity converges to 0 as $N \to \infty$ Therefore,

$$P\left\{\bigcup_{n=m}^{\infty} A_n\right\} = 1$$

which implies that $P\{A_n\ i.o.\} = 1$ and completes the proof. $\square$

§4 Applications of the Borel–Cantelli Lemmas.

In Chapter II, §3, we had several applications of the First Borel–Cantelli Lemma. In the next section we will have several more applications of this and of the Second Borel–Cantelli. Before that, we give a simple application to a fairly weak version of the strong law of large numbers.

Theorem 4.1. *Let* $\{X_i\}$ *be i.i.d. with* $EX_1 = \mu$ *and* $EX_1^4 = C\infty$*. Then* $\dfrac{S_n}{n} \to \mu$ *a.s.*

Proof. By considering $X_i' = X_i - \mu$ we may assume that $\mu = 0$. Then

$$\begin{aligned} E(S_n^4) &= E(\sum_{i=1}^{n} X_i)^4 \\ &= E \sum_{1 \le i,j,k,l \le n} X_i X_j X_k X_l \\ &= \sum_{1 \le i,j,k,l \le n} E(X_i X_j X_k X_l) \end{aligned}$$

Since the random variables have zero expectation and they are independent, the only terms in this sum which are not zero are those where all the indices are equal and those where to pair of indices are equal. That is, terms of the form EX_j^4 and $EX_i^2 X_j^2 = (EX_i^2)^2$. There are n of the first type and $3n(n-1)$ of the second type. Thus,

$$\begin{aligned} E(S_n^4) &= nE(X_1) + 3n(n-1)(EX_1^2)^2 \\ &\le Cn^2. \end{aligned}$$

By Chebyshev's inequality with $p = 4$,

$$P\{|S_n| > n\varepsilon\} \le \frac{Cn^2}{n^4 \varepsilon^4} = \frac{C}{n^2 \varepsilon^4}$$

and therefore,

$$\sum_{n=1}^{\infty} P\{|\frac{S_n}{n}| > \varepsilon\} < \infty$$

and the First Borel–Cantelli gives

$$P\left\{ \left| \frac{S_n}{n} \right| > \varepsilon \ i.o. \right\} = 0$$

which proves the result. $\square$

The following is an applications of the Second Borel–Cantelli Lemma.

Theorem 4.2. *If $X_1, X_2, \dots, X_n$ are i.i.d. with $E|X_1| = \infty$. Then $P\{|X_n| \geq n \text{ i.o.}\} = 1$ and $P\{\lim \frac{S_n}{n}$ exists $\in (-\infty, \infty)\} = 0$.*

Thus $E|X_1| < \infty$ is necessary for the strong law of large numbers. It is also sufficient.

Proof. We first note that

$$\sum_{n=1}^{\infty} P\{|X_1| \geq n\} \leq E|X_1| \leq 1 + \sum_{n=1}^{\infty} P\{|X_1| > n\}$$

which follows from the fact that

$$E|X_1| = \int_0^{\infty} P\{|X_1| > X\}dx$$

and

$$\begin{aligned} \sum_{n=0}^{\infty} \int_n^{n+1} P\{|X_1| > x\}dx &\leq \int_0^{\infty} P\{|X_1| > X\}dx \\ &\leq 1 + \sum_{n=1}^{\infty} P\{|X_1| > n\} \end{aligned}$$

Thus,

$$\sum_{n=1}^{\infty} P\{|X_n| \geq n\} = \infty$$

and therefore by the Second Borel–Cantelli Lemma,

$$P\{|X_n| > n \text{ i.o.}\} = 1.$$

Next, set

$$A = \left\{ \lim_{n\to\infty} \frac{S_n}{n} \text{ exits } \in (-\infty, \infty) \right\}$$

Clearly for $\omega \in A$,

$$\lim_{n\to\infty}\left|\frac{S_n(\omega)}{n} - \frac{S_{n+1}}{n+1}(\omega)\right| = 0$$

and

$$\lim_{n\to\infty}\frac{S_n(\omega)}{n(n+1)} = 0.$$

Hence there is an N such that for all $n > N$,

$$\lim_{n\to\infty}\left|\frac{S_n(\omega)}{n(n+1)}\right| < \frac{1}{2}.$$

Thus for $\omega \in A \cap \{\omega\colon |X_n| \geq n\ i.o.\}$,

$$\left|\frac{S_n(\omega)}{n(n+1)} - \frac{X_{n+1}(\omega)}{n+1}\right| > \frac{1}{2},$$

infinitely often. However, since

$$\frac{S_n}{n} - \frac{S_{n+1}}{n+1} = \frac{S_n}{n(n+1)} - \frac{X_{n+1}}{n+1}$$

and the left hand side goes to zero as observe above, we see that $A \cap \{\omega\colon |X_n| \geq n\ i.o.\} = \emptyset$ and since $P\{|X_n| > 1 \text{ i.o.}\} = 1$ we conclude that $P\{A\} = 0$, which completes the proof. $\square$

The following results is stronger than the Second–Borel Cantelli but it follows from it.

Theorem 4.3. *Suppose the sequence of events A_j are pairwise independent and $\sum_{j=1}^{\infty} P(A_j) = \infty$. Then*

$$\lim_{n\to\infty}\left(\frac{\sum_{j=1}^n 1_{A_j}}{\sum_{j=1}^n P(A_j)}\right) = 1.\ a.s.$$

In particular,

$$\lim_{n\to\infty}\sum_{j=1}^n 1_{A_j}(\omega) = \infty\ a.s$$

which means that

$$P\{A_n \ i.o.\} = 1.$$

Proof. Let $X_j = 1_{A_j}$ and consider their partial sums, $S_n = \sum_{j=1}^{n} X_j$. Since these random variables are pairwise independent we have as before, $\mathrm{var}(S_n) = \mathrm{var}(X_1) + \ldots + \mathrm{var}(X_n)$. Also, $\mathrm{var}(X_j) = E(X_j - EX_j|^2 \leq E(X_j)^2 = E(X_j) = P\{A_j\}$. Thus $\mathrm{var}(S_n) \leq ES_n$. Let $\varepsilon > 0$.

$$\begin{aligned} P\left\{\left|\frac{S_n}{ES_n} - 1\right| > \varepsilon\right\} &= P\left\{\left|S_n - ES_n\right| > \varepsilon ES_n\right\} \\ &\leq \frac{1}{\varepsilon^2 (ES_n)^2} \ \mathrm{var}(S_n) \\ &= \frac{1}{\varepsilon^2 ES_n} \end{aligned}$$

and this last goes to ∞ as $n \to \infty$. From this we conclude that $\dfrac{S_n}{ES_n} \to 1$ in probability. However, we have claimed a.s.

Let

$$n_k = \inf\{n \geq 1 \colon ES_n \geq k^2\}.$$

and set $T_k = S_{n_k}$. Since $EX_n \leq 1$ for all n we see that

$$k^2 \leq ET_k \leq E(T_{k-1}) + 1 \leq k^2 + 1$$

for all k. Replacing n with n_k in the above argument for S_n we get

$$\begin{aligned} P\{|T_k - ET_k| > \varepsilon ET_k\} &\leq \frac{1}{\varepsilon^2 ET_k} \\ &\leq \frac{1}{\varepsilon^2 k^2} \end{aligned}$$

Thus

$$\sum_{k=1}^{\infty} P\left\{\left|\frac{T_k}{ET_k} - 1\right| > \delta\right\} < \infty$$

and the first Borel–Cantelli gives

$$P\left\{\left|\frac{T_k}{ET_k}-1\right|>\varepsilon \ i.o\right\}=0$$

That is,

$$\frac{T_k}{ET_k}\to 1, \quad \text{a.s.}$$

Let Ω_0 with $P(\Omega_0)=1$ be such that

$$\frac{T_k(\omega)}{ET_k}\to 1.$$

for every $\omega\in\Omega_0$. Let n be any integer with $n_k\le n<n_{k+1}$. Then

$$\frac{T_k(\omega)}{ET_{k+1}}\le\frac{S_n(\omega)}{E(S_n)}\le\frac{T_{k+1}(\omega)}{ET_k}.$$

We will be done if we can show that

$$\lim_{n\to\infty}\frac{T_k(\omega)}{ET_{k+1}}\to 1 \text{ and } \lim_{n\to\infty}\frac{T_{k+1}(\omega)}{ET_k}=1.$$

Now, clearly we also have

$$\frac{ET_k}{ET_{k+1}}\cdot\frac{T_k(\omega)}{ET_k}\le\frac{S_n(\omega)}{ES_n}\le\frac{T_{k+1}(\omega)}{ET_{k+1}}\cdot\frac{ET_{k+1}}{ET_k}$$

and since

$$k^2\le ET_k\le ET_{k+1}\le(k+1)^2+1$$

we see that $1\le\frac{ET_{k+1}}{ET_k}$ and that $\frac{ET_{k+1}}{ET_k}\to 1$ and similarly $1\ge\frac{ET_k}{ET_{k+1}}$ and that $\frac{ET_k}{ET_{k+1}}\to 1$, proving the result. $\square$

§5. Convergence of Random Series, Strong Law of Large Numbers.

Definition 5.1. Let $\{X_n\}$ be a sequence of random variables. Define the σ–algebras $\mathcal{F}'_n=\sigma(X_n,X_{n+1},\dots)$ and $\mathcal{I}=\bigcap_{n\ge1}\mathcal{F}'_n$. $\mathcal{F}'_n$ is often called the "future" σ–algebra and $\mathcal{I}$ the remote (or tail) σ–algebra.

Example 5.1.

(i) If $B_n \in \mathcal{B}(\mathbb{R})$, then $\{X_n \in B_n \text{ i.o.}\} \in \mathcal{I}$. and if we take $X_n = 1_{A_n}$ we see that Then $\{A_n \text{ i.o.}\} \in \mathcal{I}$.

(ii) If $S_n = X_1 + \ldots + X_n$, then clearly, $\{\lim_{n\to\infty} S_n \text{ exists}\} \in \mathcal{I}$ and $\{\overline{\lim} S_n/c_n > \lambda\} \in \mathcal{I}$, if $c_n \to \infty$. However, $\{\overline{\lim} S_n > 0\} \notin \mathcal{I}$

Theorem 5.1 (Kolmogorov $0-1$ Law). *If $X_1, X_2 \ldots$ are independent and $A \in \mathcal{I}$, then $P(A) = 0$ or 1.*

Proof. We shall show that A is "independent of itself" and hence $P(A) = P(A \cap A) = P(A)P(A)$ which implies that $P(A) = 0$ or 1. First, since $X_1, \ldots$ are independent if $A \in \sigma(X_1, \ldots, X_n)$ and $B \in \sigma(X_{n+1}, \ldots)$ then A and B are independent. Thus if, $A \in \sigma(X_1, \ldots, X_n)$ and $B \in \mathcal{I}$, then A and B are independent. Thus $\bigcup\limits_n \sigma(X_1, \ldots, X_n)$ is independent of $\mathcal{I}$. Since they are both π–systems (clearly if $A, B \in \bigcup\limits_n \sigma(X_1, \ldots, X_n)$ then $A \in \sigma(X_1 \ldots, X_n)$ and $B \in \sigma(X_1, \ldots, X_m)$ for some n and m and so $A \cap B \in \sigma(X_1, \ldots, X_{\max(n,m)})$), $\bigcup\limits_n \sigma(X_1, \ldots, X_n)$ is independent of $\mathcal{I}$, by Theorem 2.3, Chapter IV. Since $A \in \mathcal{I}$ implies $A \in \sigma(X_1, X_2, \ldots)$, we are done. $\square$

Corollary. *Let A_n be independent. Then $P\{A_n \text{ i.o.}\} = 0$ or 1. In the same way, if $X - n$ are independent then $P\{\lim S_n \text{ exists}\} = 0$ or1.*

Or next task is to investigate when the above probabilities are indeed one. Recall that Chebyshev's inequality gives, for mean zero random variables which are independent, that

$$P\{|S_n| > \lambda\} \leq \frac{1}{\lambda^2} \operatorname{var}(S_n).$$

The following results is stronger and more useful as we shall see soon.

Theorem 5.2 (Kolmogorov's inequality). *Suppose $\{X_n\}$ be independent, $EX_n = 0$ and $var(X_n) < \infty$ for all n. Then for any $\lambda > 0$,*

$$\begin{aligned} P\{\max_{1 \le k \le n} |S_k| \ge \lambda\} &\le \frac{1}{\lambda^2} E|S_n|^2 \\ &= \frac{1}{\lambda^2}\ var(S_n). \end{aligned}$$

Proof. Set

$$A_k = \{\omega \subset \Omega\colon |S_k(\omega)| \ge \lambda, |S_j(\omega)| < \lambda \text{ for all } j < k\}$$

Note that these sets are disjoint and

$$\begin{aligned} ES_n^2 &\ge \sum_{k=1}^n \int_{A_k} S_n^2 dP = \\ &= \sum_{k=1}^n \int_{A_k} S_k^2 + 2S_k S_n - 2S_k + (S_n - S_k)^2 dP \\ &\ge \sum_{k=1}^n \int_{A_k} S_k^2 dP + 2\sum_{k=1}^n \int_\Omega S_k 1_{A_k}(S_n - S_k) dP \end{aligned} \tag{5.1}$$

Now,

$$S_k 1_{A_k} \in \sigma(X_1, \ldots, X_k)$$

and

$$S_n - S_k \in \sigma(X_{k+1} \ldots S_n)$$

and hence they are independent. Since $E(S_n - S_k) = 0$, we have $E(S_k 1_{A_k}(S_n - S_k)) = 0$ and therefore the second term in (5.1) is zero and we see that

$$\begin{aligned} E{S_n}^2 &\ge \sum_{k=1}^n \int_{A_k} |S_k^2| dp \\ &\ge \lambda^2 \sum_{k=1}^\infty P(A_k) \\ &= \lambda^2 P(\max_{1 \le k \le n} |S_k| > \lambda\Big\} \end{aligned}$$

which proves the theorem. $\square$

Theorem 5.3. *If X_j are independent, $EX_j = 0$ and $\sum_{n=1}^{\infty} var(X_n) < \infty$, then $\sum_{n=1}^{\infty} X_n$ converges a.s.*

Proof. By Theorem 5.2, for $N > M$ we have

$$P\Big\{\max_{M \le n \le N} |S_n - S_M| > \varepsilon\Big\} \le \frac{1}{\varepsilon^2}\ \mathrm{var}(S_N - S_M)$$
$$= \frac{1}{\varepsilon^2} \sum_{n=M+1}^{N} \mathrm{var}(X_n).$$

Letting $N \to \infty$ gives $P\{\max_{m \ge M} |S_n - S_M| > \varepsilon\} \le \frac{1}{\varepsilon^2} \sum_{n=M+1}^{\infty} \varepsilon(X_n)$ and this last quantity goes to zero as $M \to \infty$ since the sum converges. Thus if

$$\Lambda_M = \sup_{n,m \ge M} |S_m - S_n|$$

then

$$P\{\Lambda_M > 2\varepsilon\} \le P\{\max_{m \le M} |S_m - S_M| > \varepsilon\} \to 0$$

as $M \to \infty$ and hence $\Lambda_M \to 0$ a.s. as $M \to \infty$. Thus for almost every ω, $\{S_m(\omega)\}$ is a Cauchy sequence and hence it converges. $\square$

Example 5.2. Let $X_1, X_2, \ldots$ be i.i.d. $N(0,1)$. Then for every t,

$$B_t(\omega) = \sum_{n=1}^{\infty} X_n \frac{\sin(n\pi t)}{n}$$

converges a.s. (This is a series representation of Brownian motion.)

Theorem 5.4 (Kolmogorov's Three Series Theorem). *Let $\{X_j\}$ be independent random variables. Let $A > 0$ and set $Y_j = X_j 1_{(|X_j| \le A)}$. Then $\sum_{n=1}^{\infty} X_n$ converges a.s. if and only if the following three conditions hold:*

(i) $\sum_{n=1}^{\infty} P(|X_n| > A) < \infty$, *(ii)* $\sum_{n=1}^{\infty} EY_n$ *converges,*

(iii) $\sum_{n=1}^{\infty} var(Y_n) < \infty.$

Proof. Assume (i)–(iii). Let $\mu_n = EY_n$. By (iii) and Theorem 5.3, $\sum(Y_n - \mu_n)$ converges a.e. This and (ii) show that $\sum_{n=1}^{\infty} Y_n$ converges a.s. However, (i) is equivalent to $\sum_{n=1}^{\infty} P(X_n \neq Y_n) < \infty$ and by the Borel–Cantelli Lemma,

$$P(X_n \neq Y_n \text{ i.o.}\} = 0.$$

Therefore, $P(X_n = Y_n \text{ eventually}\} = 1$. Thus if $\sum_{n=1}^{\infty} Y_n$ converges, so does $\sum_{n=1}^{\infty} X_n$.

We will prove the necessity later as an application of the central limit theorem. □

For the proof of the strong law of large numbers, we need

Lemma 5.1 (Kronecker's Lemma). *Suppose that a_n is a sequence of positive real numbers converging up to ∞ and suppose $\sum_{n=1}^{\infty} \frac{x_n}{a_n}$ converges. Then*

$$\frac{1}{a_n} \sum_{m=1}^{n} x_m \to 0.$$

Proof. Let $b_n = \sum_{j=1}^{n} \frac{x_j}{a_j}$. Then $b_n \to b_\infty$, by assumption. Set $a_0 = 0,\ b_0 = 0$. Then

$x_n = a_n(b_n - b_{n-1}),\ n = 1, 2, \ldots$ and

$$\begin{aligned}
\frac{1}{a_n}\sum_{j=1}^{n} x_j &= \frac{1}{a_n}\sum_{j=1}^{n} a_j(b_j - b_{j-1}) \\
&= \frac{1}{a_n}\left[b_n a_n - b_0 a_0 - \sum_{j=0}^{n-1} b_j(a_{j+1} - a_j)\right] \\
&= b_n - \frac{1}{a_n}\sum_{j=0}^{n-1} b_j(a_{j+1} - a_j)
\end{aligned}$$

The last equality is by summation by parts. To see this, precede by induction observing first that

$$\begin{aligned}
\sum_{j=1}^{n} a_j(b_j - b_{j-1}) &= \sum_{j=1}^{n-1} a_j(b_j - b_{j-1}) + a_n(b_n - b_{n-1}) \\
&= b_{n-1}a_{n-1} - b_0 a_0 - \sum_{j=0}^{n-2} b_j(a_{j+1} - a_j) + a_n b_n - a_n b_{n-1} \\
&= a_n b_n - b_0 a_0 - \sum_{j=0}^{n-2} b_j(a_{j+1} - a_j) - b_{n-1}(a_n - a_{n-1}) \\
&= a_n b_n - b_0 a_0 - \sum_{j=0}^{n-1} b_j(a_{j+1} - a_j)
\end{aligned}$$

Now, recall a $b_n \to b_\infty$. We claim that $\dfrac{1}{a_n}\displaystyle\sum_{j=0}^{n-1} b_j(a_{j+1} - a_j) \to b_\infty$. Since $b_n \to b_\infty$, given $\varepsilon > 0\ \exists\ N$ such that for all $j > N$, $|b_j - b_\infty| < \varepsilon$. Since

$$\frac{1}{a_n}\sum_{j=0}^{n-1}(a_{j+1} - a_j) = 1$$

Jensen's inequality gives

$$
\begin{aligned}
|\frac{1}{a_n}\sum_{j=0}^{n-1} b_j(a_{j+1}-a_j) - b_\infty| &\le \left|\frac{1}{a_n}\sum_{j=1}^{n-1}(b_\infty - b_j)(a_{j+1}-a_j)\right| \\
&\le \frac{1}{a_n}\sum_{j=1}^{N}|(b_\infty - b_j)(a_{j+1}-a_j)| \\
&+ \frac{1}{a_n}\sum_{j=N+1}^{n-1}|b_\infty - b_j||a_{j+1}-a_j| \\
&\le \frac{1}{a_n}\sum_{j=1}^{N}|b_N - b_j||a_{j+1}-a_j| + \varepsilon\frac{1}{a_n}\sum_{j=N+1}^{n}|a_{j+1}-a_j| \\
&\le \frac{M}{a_n} + \varepsilon.
\end{aligned}
$$

Letting first $n \to \infty$ and then $\varepsilon \to 0$ completes the proof. $\square$

Theorem 5.5 (The strong law of large numbers). *Suppose* $\{X_j\}$ *are i.i.d.,* $E|X_1| < \infty$ *and set* $EX_1 = \mu$. *Then* $\frac{S_n}{n} \to \mu$ *a.s.*

Proof. Let $Y_k = X_k 1_{(|X_k| \le k)}$ Then

$$
\begin{aligned}
\sum_{n=1}^{\infty} P\{X_k \ne Y_k\} &= \sum P(|X_k| > k\} \\
&\le \int_0^\infty P(|X_1| > \lambda\} d_\lambda \\
&= E|X| < \infty.
\end{aligned}
$$

Therefore by the First Borel–Cantelli Lemma, $P\{X_k \ne Y_k \ i.o.\} = 0$ or put in other words, $P\{X_k \ne Y_k \text{ eventually}\} = 1$. Thus if we set $T_n = Y_1 + \ldots + Y_n$. It suffices to prove $\frac{T_n}{n} \to \mu$ a.s. Now set $Z_k = Y_k - EY_k$. Then $E(Z_k) = 0$ and

$$\begin{aligned}
\sum_{k=1}^{\infty} \frac{\operatorname{var}(Z_n)}{k^2} &\leq \sum_{k=1}^{\infty} \frac{E(Y_k^2)}{k^2} \\
&= \sum_{k=1}^{\infty} \frac{1}{k^2} \int_0^{\infty} 2\lambda P\{|Y_k| > \lambda\} d\lambda \\
&= \sum_{k=1}^{\infty} \frac{1}{k^2} \int_0^{k} 2\lambda P\{|X_1| > \lambda\} d\lambda \\
&= 2 \int_0^{\infty} \sum_{k=1}^{\infty} \frac{\lambda}{k^2} 1_{\{\lambda \leq k\}}(\lambda) P\{|X_1| > \lambda\} d\lambda \\
&= 2 \int_0^{\infty} \lambda \left(\sum_{k>\lambda} \frac{1}{k^2} \right) P\{|X_1| > \lambda\} d\lambda. \\
&\leq CE|X_1|,
\end{aligned}$$

where we used the fact that

$$\sum_{k>\lambda} \frac{1}{k^2} \leq \frac{C}{\lambda}$$

for some constant C which follows from the integral test. By Theorem 5.3, $\sum_{k=1}^{\infty} \frac{Z_k}{k}$ converges a.s. and the Kronecker's Lemma gives that

$$\frac{1}{n} \sum_{k=1}^{n} Z_k \to 0 \ a.s.$$

which is the same as

$$\frac{1}{n} \sum_{k=1}^{n} (Y_k - EY_k) \to 0 \text{ a.s.}$$

or

$$\frac{T_n}{n} - \frac{1}{n} \sum_{k=1}^{n} EY_k \to 0 \text{ a.s.}$$

We would be done if we can show that

$$\frac{1}{n} \sum_{k=1}^{n} EY_k \to \mu. \tag{5.2}$$

We know $EY_k \to \mu$ as $k \to \infty$. That is, there exists an N such that for all $k > N$, $|EY_k - \mu| < \varepsilon$. With this N fixed we have for all $n \geq N$,

$$\begin{aligned}\left|\frac{1}{n}\sum_{k=1}^{n} EY_k - \mu\right| &= \left|\frac{1}{n}\sum_{k=1}^{n}(EY_k - \mu)\right| \\ &\leq \frac{1}{n}\sum_{k=1}^{N} E|Y_k - \mu| + \frac{1}{n}\sum_{k=N}^{n} E|Y_k - \mu| \\ &\leq \frac{1}{n}\sum_{k=1}^{N} E|Y_k - \mu| + \varepsilon.\end{aligned}$$

Let $n \to \infty$ to complete the proof. $\square$

§6. Variants of the Strong Law of Large Numbers.

Let us assume $E(X_i) = 0$ then under the assumptions of the strong law of large numbers we have $\frac{S_n}{n} \to 0$ a.s. The question we address now is: Can we have a better rate of convergence? The answer is yes under the right assumptions and we begin with

Theorem 6.1. *Let $X_1, X_2, \dots$ be i.i.d., $EX_i = 0$ and $EX_1^2 \leq \sigma^2 < \infty$. Then for any $\varepsilon \geq 0$*

$$\lim_{n\to\infty} \frac{S_n}{n^{1/2}(\log n)^{1/2+\varepsilon}} = 0,$$

a.s.

We will show later that in fact,

$$\overline{\lim} \frac{S_n}{\sqrt{2\sigma^2 n \log\log n}} = 1,$$

a.s. This last is the celebrated law of the iterated logarithm of Kinchine.

Proof. Set $a_n = \sqrt{n}(\log n)^{\frac{1}{2}+\varepsilon}$, $n \geq 2$. $a_1 > 0$

$$\sum_{n=1}^{\infty} \operatorname{var}(X_n/a_n) = \sigma\left(\frac{1}{a_1^2} + \sum_{n=2}^{\infty} \frac{1}{n(\log n)^{1+2\varepsilon}}\right) < \infty.$$

Then $\sum_{n=1}^{\infty} \frac{X_n}{a_n}$ converges a.s. and hence $\frac{1}{a_n} \sum_{k=1}^{n} X_k \to 0$ a.s. $\square$

What if $E|X_1|^2 = \infty$ But $E|X_1|^p < \infty$ for some $1 < p < 2$? For this we have

Theorem 6.2 (Marcinkiewidz and Zygmund). *X_j i.i.d., $EX_1 = 0$ and $E|X_1|^p < \infty$ for some $1 < p < 2$. Then*

$$\lim_{n\to\infty} \frac{S_n}{n^{1/p}} = 0,$$

a.s.

Proof. Let

$$Y_k = X_k 1_{(|X_n| \leq k^{1/p})}$$

and set

$$T_n = \sum_{k=1}^{n} Y_k.$$

It is enough to prove, as above, that $\frac{T_n}{n^{1/p} \to 0}$, a.s. To see this, observe that

$$\begin{aligned} \sum P\{Y_k \neq X_k\} &= \sum_{k=1}^{\infty} P\{|X_k|^p > k\} \\ &\leq E(|X_1|^p) < \infty \end{aligned}$$

and therefore by the first Borel–Cantelli Lemma, $P\{Y_k \neq X_k \text{ i.o.}\} = 0$ which is the same as $P\{Y_k = X_k, \text{ eventually}\} = 1$

Next, estimating by the integral we have

$$\begin{aligned} \sum_{k > \lambda^p} \frac{1}{k^{2/p}} &\leq C \int_{\lambda^p}^{\infty} \frac{dx}{x^{2/p}} \\ &= \frac{1}{(1-2/p)} x^{2-2/p} \Big|_{\lambda^p}^{\infty} \\ &= \lambda^{p-2} \end{aligned}$$

and hence

$$\begin{aligned}
\sum_{k=1}^{\infty} \operatorname{var}(Y_k/k^{1/p}) &\le \sum_{k=1}^{\infty} \frac{EY_k^2}{k^{2/p}} \\
&= 2\sum_{k=1}^{\infty} \frac{1}{k^{2/p}} \int_0^{\infty} \lambda P\{|Y_k| > \lambda\} d\lambda \\
&= 2\sum_{k=1}^{\infty} \frac{1}{k^{2/p}} \int_0^{k^{1/p}} \lambda P\{|X_1| > \lambda\} d\lambda \\
&= 2\sum_{k=1}^{\infty} \frac{1}{k^{2/p}} \int_0^{\infty} 1_{(0,k^{1/p})}(\lambda) \lambda P\{|X_1| > \lambda\} d\lambda \\
&= 2\int_0^{\infty} \lambda P\{|X_1| > \lambda\} \left(\sum_{k > \lambda^p} \frac{1}{k^{2/p}} \right) d\lambda \\
&\le 2\int_0^{\infty} \lambda^{p-1} P\{|X_1| > \lambda\} d\lambda = C_p E|X_1|^p < \infty.
\end{aligned}$$

Thus, and Kronecker implies, with $\mu_k = E(Y_k)$, that

$$\frac{1}{n^{1/p}} \sum_{k=1}^{n} (Y_k - \mu_k) \to 0, \quad \text{a.s.}$$

If we are bale to show that $\frac{1}{n^{1/p}} \sum_{k=1}^{n} \mu_k \to 0$, we will be done. Observe that $0 = E(X_1) = E(X1_{(|X| \ge k^{1/p})}) + \mu_k$ so that $|\mu_k| \le |E(X1_{(|X| \ge k^{1/p})}|$ and therefore

$$\begin{aligned}
\left| \frac{1}{n^{1/p}} \sum_{k=1}^{n} \mu_k \right| &\le \frac{1}{n^{1/p}} \sum_{k=1}^{n} \int_{k^{1/p}}^{\infty} P\{|X_1| > \lambda\} d\lambda \\
&\le \frac{1}{pn^{1/p}} \sum_{k=1}^{n} \frac{1}{k^{1-1/p}} \int_{k^{1/p}}^{\infty} p\lambda^{p-1} P\{|X_1| > \lambda\} d\lambda \\
&= \frac{1}{pn^{1/p}} \sum_{k=1}^{n} \frac{1}{k^{1-1/p}} E\{|X_1|^p; |X_1| > k^{1/p}\}.
\end{aligned}$$

Since $X \in L^p$, given $\varepsilon > 0$ there is an N such that $E(|X_1|^p |X_1| > k^{1/p}) < \varepsilon$ if $k > N$. Also,

$$\sum_{k=1}^{n} \frac{1}{k^{1-1/p}} \le C \int_1^n x^{1/p-1} dx \le Cn^{1/p}.$$

The Theorem follows from these. $\square$

Theorem 6.3. *Let $X_1, X_2, \dots$ be i.i.d. with $EX_j^+ = \infty$ and $EX_j^- < \infty$. Then*

$$\lim_{n\to\infty} \frac{S_n}{n} = \infty, \quad a.s.$$

Proof. Let $M > 0$ and $X_j^M = X_i \wedge M$, the maximum of X_j and M. Then X_i^M are i.i.d. and $E|X_i^M| < \infty$. (Here we have used the fact that $EX_j^- < \infty$.) Setting $S_n^M = X_1^M = X_1^M + \dots + X_n^M$ we see that $\dfrac{S_n^M}{n} \to EX_1^M$ a.s. Now, since $X_i \geq X_i^M$ we have

$$\underline{\lim}\ \frac{S_n}{n} \geq \lim_{n\to\infty} \frac{S_n^M}{n} = EX_1^M, \ a.s.$$

However, by the monotone convergence theorem, $E(X_1^M)^+ \uparrow E(X_1^+) = \infty$, hence

$$EX_1^M = E(X_1^M)^+ - E(X_1^M)^- \uparrow +\infty.$$

Therefore,

$$\underline{\lim} \frac{S_n}{n} = \infty, \ a.s.$$

and the result is proved. $\square$

§7. Two Applications.

We begin with an example from Renewal Theory. Suppose $X_1, X_2, \dots$ be are i.i.d. and $0 < X_i < \infty$, a.s. Let $T_n = X_1 + \dots + X_n$ and think of T_n as the time of the nth occurrence of an event. For example, X_i could be the lifetime of ith lightbulb in a room with infinitely many lightbulbs. Then $T_n =$ is the time the nth lightbulb burns out. Let $N_t = \sup\{n: T_n \leq t\}$ which in this example is the number of lightbulbs that have burnt out by time t.

Theorem 7.1. *Let X_j be i.i.d. and set $EX_1 = \mu$ which may or may not be finite. Then $\frac{N_t}{t} \to 1/\mu$, a.s. as $t \to \infty$ where this is 0 if $\mu = \infty$. Also, $E(N(t))/t \to 1/\mu$*

Continuing with our lightbulbs example, note that if the mean lifetime is large then the number of lightbulbs burnt by time t is very small.

Proof. We know $\dfrac{T_n}{n} \to \mu$ a.s. Note that for every $\omega \in \Omega,\ N_t(\omega)$ is and integer and

$$T(N_t) \leq t < T(N_t + 1).$$

Thus,

$$\frac{T(N_t)}{N_t} \leq \frac{t}{N_t} \leq \frac{T(N_t+1)}{N_t+1} \cdot \frac{N_t+1}{N_t}.$$

Now, since $T_n < \infty$ for all n, we have $N_t \uparrow \infty$ a.s. By the law of large numbers there is an $\Omega_0 \subset \Omega$ such that $P(\Omega_0) = 1$ and such that for $\omega \in \Omega_0$,

$$\frac{T_{N_t(\omega)}(\omega)}{N_t(\omega)} \to \mu \text{ and } \frac{N_t(\omega)+1}{N_t(\omega)} \to 1.$$

Thus $t/N_t(\omega) \to \mu$ a.s. and we are done. $\square$

Let $X_1, X_2, \dots$ be i.i.d. with distribution F. For $x \in \mathbb{R}$ set

$$F_n(x,\omega) = \frac{1}{n} \sum_{n=1}^{n} 1_{(X_k \leq x)}(\omega).$$

This is the observed frequency of values $\leq x$. Now, fix $\omega \in \Omega$ and set $a_k = X_k(\omega)$. Then $F_n(x,\omega)$ is the distribution with jump of size $\frac{1}{n}$ at the points a_k. This is called the *imperial distribution based on n samples of F*. On the other hand, let us fix x. Then $F_n(x,\cdot)$ is a random variable. What kind of a random variable is it? Define

$$\rho_k(\omega) = 1_{(X_k \leq x)}(\omega) = \begin{cases} 1, & X_k(\omega) \leq x \\ 0, & X_k(\omega) > x \end{cases}$$

Notice that in fact the ρ_k are independent Bernoullians with $p = F(x)$ and $E\rho_k = F(x)$. Writing

$$F_n(x,\omega) = \frac{1}{n} \sum_{k=1}^{n} \rho_k$$

we see that in fact $F_n(x,\cdot) = \frac{S_n}{n}$ and the Strong Law of Large numbers shows that for every $x \in \mathbb{R}$, $F_n(x,\omega) \to F(x)$ a.s. Of course, the exceptional set may depend on x. That is, what we have proved here is that given $x \in \mathbb{R}$ there is a set

$N_x \subset \Omega$ such that $P(N_x) = 0$ and such that $F_n(x, \omega) \to F(x)$ for $\omega \in N_x$. If we set $N = \cup_{x \in Q} N_x$ where we use Q to denote the rational numbers, then this set also has probability zero and off this set we have $F_n(x, \omega) \to F(x)$ for all $\omega \in N$ and all $x \in Q$. This and the fact that the discontinuities of distribution functions are at most countable turns out to be enough to prove

Theorem 7.2 (Glivenko–Cantelli Theorem). *Let*

$$D_n(\omega) = \sup_{x \in \mathbb{R}} |F_n(x, \omega) - F(x)|.$$

Then $D_n \to 0$ *a.s.*

VI
THE CENTRAL LIMIT THEOREM

§1 Convergence in Distribution.

If X_n "tends" to a limit, what can you say about the sequence F_n of d.f. or the sequence $\{\mu_n\}$ of measures?

Example 1.1. *Suppose X has distribution F and define the sequence of random variables $X_n = X+1/n$. Clearly, $X_n \to X$ a.s. and in several other ways. $F_n(x) = P(X_n \le x) = P(X \le x - 1/n) = F(x - 1/n)$. Therefore,*

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

Hence we do not have convergence of F_n to F. Even worse, set $X_n = X + C_n$ where $C_n = \begin{cases} \frac{1}{n} & even \\ -1/n & odd \end{cases}$. Thenthe limit may not even exist.

Definition 1.1. *The sequence $\{F_n\}$ of d.f. converges weakly to the d.f. F if $F_n(x) \to F(x)$ for every point of continuity of F. We write $F_n \Rightarrow F$. In all our discussions we assume F is a d.f. but it could just as well be a (sub. d.f.).*

The sequence of random variables X_n converges weakly to X if their distributions functions $F_n(x) = P(X_n \le x)$ converge weakly to $F(x) = P(X \le x)$. We will also use $X_n \Rightarrow X$.

Example 1.2.

(1) The Glivenko–Cantelli Theorem

(2) $X_i =$ *i.i.d.* ± 1, *probability* $1/2$. *If* $S_n = X_1 + \ldots + X_n$ *then*

$$F_n(y) = P\left(\frac{S_n}{\sqrt{n}} \le y\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-x^2/2} dx.$$

This last example can be written as $\frac{S_n}{n} \Rightarrow N(0,1)$ and is called the *the De Moivre–Laplace Central limit Theorem.* Our goal in this chapter is to obtain a very general version of this result. We begin with a detailed study of convergence in distribution.

Theorem 1.1 (Skorhod's Theorem). *IF* $F_n \Rightarrow F$, *then there exists random variables* Y_n, Y *with* $Y_n \to Y$ *a.s. and* $Y_n \sim F_n$, $Y \sim F$.

Proof. We construct the random variables on the canonical space. That is, let $\Omega = (0,1)$, $\mathcal{F}$ the Borel sets and P the Lebesgue measure. As in Chapter IV, Theorem 1.1,

$$Y_n(\omega) = \inf\{x\colon \omega \le F_n(x)\},\ Y(\omega) = \inf\{x\colon \omega \le F(x)\}.$$

are random variables satisfying $Y_n \sim F_n$ and $Y \sim F$

The idea is that if $F_n \to F$ then $F_n^{-1} \to F^{-1}$, but of course, the problem is that this does not happen for every point and that the random variables are not exactly inverses of the distribution functions. Thus, we need to proceed with some care. In fact, what we shall show is that $Y_n(\omega) \to Y(\omega)$ except for a countable set. Let $0 < \omega < 1$. Given $\varepsilon > 0$ chose and x for which $Y(\omega) - \varepsilon < x < Y(\omega)$ and $F(x-) = F(x)$, (that is for which F is continuous at x). Then by definition $F(x) < \omega$. Since $F_n(x) \to F(x)$ we have that for all $n > N$, $F_n(x) < \omega$. Hence, again by definition, $Y(\omega) - \varepsilon < x < Y_n(\omega)$, for all such n. Therefore,

$$\underline{\lim}\, Y_n(\omega) \ge Y(\omega).$$

It remains to show that

$$\overline{\lim}\, Y_n(x) \le Y(x).$$

Now, if $\omega < \omega'$ and $\varepsilon > 0$, choose y for which $Y(\omega') < y < Y(\omega') + \varepsilon$ and F is continuous at y. Now,

$$\omega < \omega' \leq F(Y(\omega')) \leq F(y).$$

Again, since $F_n(y) \to F(y)$ we see that for all $n > N$, $\omega \leq F_n(y)$ and hence $Y_n(\omega) \leq y < Y(\omega') + \varepsilon$ which implies $\overline{\lim}\, Y_n(\omega) \leq Y(\omega')$. If Y is continuous at ω we must have

$$\overline{\lim}\, Y_n(\omega) \leq Y(\omega).$$

□

The following corollaries follow immediately from Theorem 1.1 and the results in Chapter II.

Corollary 1.1 (Fatou's in Distribution). *Suppose $X_n \Rightarrow X$ and $g \geq 0$ and continuous. Then $E(g(X)) \leq \underline{\lim} E(g(X_n))$.*

Corollary 1.2 (Dominated Convergence in Distribution). *If $X_n \Rightarrow X$, g is continuous and and $E|(g(X_n)| < C$, then*

$$E(g(X_n)) \to E(g(X)).$$

The following is a useful characterization of convergence in distribution.

Theorem 1.2. *$X_n \Rightarrow X$ if and only if for every bounded continuous function g we have $E(g(X_n)) \to E(g(X))$.*

Proof. If $X_n \Rightarrow X$ then Corollary 2.1 implies the convergence of the expectations. Conversely, let

$$g_{x,\varepsilon}(y) = \begin{cases} 1 & y \leq x \\ 0 & y \geq x + \varepsilon \\ linear & x \leq y \leq x + \varepsilon \end{cases}$$

It follows from this that

$$P(X_n \le x) \le E(g_{x,\varepsilon}(X_n))$$

and therefore,

$$\begin{aligned}\overline{\lim} P(X_n \le x) &\le \overline{\lim} E(g_{x,\varepsilon}(X_n)) \\ &= E(g_{x,\varepsilon}(X)) \\ &\le P(X \le x + \varepsilon).\end{aligned}$$

Now, let $\varepsilon \to 0$ to conclude that

$$\overline{\lim} P(X_n \le x) \le P(X \le x).$$

In the same way,

$$\begin{aligned}P(X \le x - \varepsilon) &\le E(g_{x-\varepsilon,\varepsilon}(X)) \\ &= \underline{\lim}_{n\to\infty} E(g_{x-\varepsilon,\varepsilon}(X_n)) \\ &\le \underline{\lim}_{n\to\infty} P(X_n \le x).\end{aligned}$$

Now, let $\varepsilon \to 0$. If F continuous at x, we obtain the result. $\square$

Corollary 1.3. *Suppose $X_n \to X$ in probability. Then $X_n \Rightarrow X$.*

Lemma 1.1. *Suppose $X_n \to 0$ in probability and $|X_n| \le Y$ with $E(Y) < \infty$. Then $E|X_n| \to 0$.*

Proof. Fix $\varepsilon > 0$. Then $P\{|X_n| > \varepsilon\} \to 0$, as $n \to \infty$. Hence by Proposition 2.6 in Chapter II,

$$\int_{\{|X_n| > \varepsilon\}} |Y| \, dP \to 0, \text{as } n \to \infty.$$

Since

$$
\begin{aligned}
E|X_n| &= \int_{\{|X_n|<\varepsilon\}} |X_n| dP + \int_{\{|X_n|>\varepsilon\}} |X_n| dP \\
&< \varepsilon + \int_{\{|X_n|>\varepsilon\}} |Y| \, dP,
\end{aligned}
$$

the result follows. □

Proof of Corollary 1.3. If $X_n \to X$ in probability and g is bounded and continuous then $g(X_n) \to g(X)$ in probability (why ?) and hence $E(g(X_n)) \to E(g(X))$, proving $X_n \Rightarrow X$.

An alternative proof is as follows. Set $a_n = E(G(X_n))$ and $a = E(X)$. Let a_{n_k} be a subsequence. Since X_{n_k} converges to X in probability also, we have a subsequence $X_{n_{k_j}}$ which converges almost everywhere and hence by the dominated convergence theorem we have $a_{n_{k_j}} \to a$ and hence the sequence a_n also converges to a, proving the result. □

Theorem 1.3 (Continuous mapping Theorem). *Let g be a measurable function in $\mathbb{R}$ and let $D_g = \{x\colon g \text{ is discontinuous at } x\}$. If $X_n \Rightarrow X$ and $P\{X \in D_g\} = \mu(D_g) = 0$, then $g(X_n) \Rightarrow g(X)$*

Proof. Let $X_n \sim Y_n,\ X \sim Y$ and $Y_n \to Y$ a.s. Let f be continuous and bounded. Then $D_{f \circ g} \subset D_g$. So,

$$P\{Y_\infty \in D_{f\circ g}\} = 0.$$

Thus,

$$f(g(Y_n)) \to f(g(Y))$$

a.s. and the dominated convergence theorem implies that $E(f(g(Y_n))) \to E(f(g(Y))$ and this proves the result.

Next result gives a number of useful equivalent definitions.

Theorem 1.4. *The following are equivalent:*

(i) $X_n \Rightarrow X$.

(ii) *For all open sets* $G \subset \mathbb{R}$, $\underline{\lim} P(X_n \in G) \geq P(X \in G)$ *or what is the same,* $\underline{\lim} \mu_n(G) \geq \mu(G)$, *where* $X_n \sim \mu_n$ *and* $X \sim \mu$.

(iii) *For all closed sets* $K \subset \mathbb{R}$, $\overline{\lim} P(X_n \in K) \leq P(X \in K)$.

(iv) *For all sets* $A \subset \mathbb{R}$ *with* $P(X \in \partial A) = 0$ *we have* $\lim_{n \to \infty} P(X_n \in A) = P(X \in A)$.

We recall that for any set A, $\partial A = \overline{A} \backslash A^0$ where $\overline{A}$ is the closure of the set and A^0 is its interior. It can very well be that we have strict inequality in (ii) and (iii). Consider for example, $X_n = 1/n$ so that $P(X_n = 1/n) = 1$. Take $G = (0, 1)$. Then $P(X_n \in G) = 1$. But $1/n \to 0 \in \partial G$, so,

$$P(X \in G) = 0.$$

Also, the last property can be used to define weak convergence of probability measures. That is, let μ_n and μ be probability measures on $(\mathbb{R}, \mathcal{B})$. We shall say that μ_n converges to μ weakly if $\mu_n(A) \to \mu(A)$ for all borel sets A in $\mathbb{R}$ with the property that $\mu(\partial A) = 0$.

Proof. We shall prove that (i) $\Rightarrow$ (ii) and that (ii)$\Leftrightarrow$ (iii). Then that (ii) and (iii) $\Rightarrow$ (iv), and finally that (iv) $\Rightarrow$ (i).

Proof. Assume (i). Let $Y_n \sim X_n,\ Y \sim X,\ Y_n \to Y$ a.s. Since G is open,

$$\underline{\lim} 1_{(Y_n \in G)}(\omega) \geq 1_{(Y \in G)}(\omega)$$

Therefore Fatou's Lemma implies

$$P(Y \in G) \leq \underline{\lim} P(Y_n \in G),$$

proving (ii). Next, (ii) $\Rightarrow$ (iii). Let K be closed. Then K^c is open. Put

$$P(X_n \in K) = 1 - P(X_n \in K^c)$$
$$P(X \in K) = 1 - P(X \in K^c).$$

The equivalence of (ii) and (iii) follows from this.

Now, (ii) and (iii) $\Rightarrow$ (iv). Let $K = \overline{A}$, $G = A^0$ and $\partial A = \overline{A} \backslash A^0$. Now, $G = K \backslash \partial A$ and under our assumption that $P(X \in \partial A) = 0$,

$$P(X \in K) = P(X \in A) = P(X \in G).$$

Therefore, (ii) and (iii) $\Rightarrow$

$$\begin{aligned} \overline{\lim}\, P(X_n \in A) &\leq \overline{\lim}\, P(X_n \in K) \\ &\leq P(X \in K) \\ &= P(X \in A) \\ \underline{\lim}\, P(X_n \in A) &\geq \underline{\lim}\, P(X_n \in G) \end{aligned}$$

and this gives

$$P(X_\infty \in G) = P(X_\infty \in A).$$

To prove that (iv) implies (i), take $A = (-\infty, x]$. Then $\partial A = \{x\}$ and this completes the proof. $\square$

Next, recall that any bounded sequence of real numbers has the property that it contains a subsequence which converges. Suppose we have a sequence of probability measures μ_n. Is it possible to pull a subsequence μ_{u_k} so that it converges weakly to a probability measure μ? Or, is it true that given distribution functions F_n there is a subsequence $\{F_{n_k}\}$ such that F_{n_k} converges weakly to a distribution function F? The answer is no, in general.

Example 1.3. Take $F_n(x) = \frac{1}{3}1_{(x \geq n)}(x) + \frac{1}{3}1_{(x \geq -n)}(x) + \frac{1}{3}G(x)$ where G is a distribution function. Then

$$\begin{aligned}
&\lim_{n \to \infty} F_n(x) = F(x) = 1/3 + 1/3G(x) \\
&\lim_{x \uparrow \infty} f(x) = 2/3 < 1 \\
&\lim_{x \downarrow -\infty} F(x) = 1/3 \neq 0.
\end{aligned}$$

Lemma 1.1. *Let f be an increasing function on the rationals Q and define $\tilde{f}$ on $\mathbb{R}$ by*

$$\begin{aligned}
\tilde{f}(x) &= \inf_{x < t \in Q} f(t) = \inf\{f(t) \colon x < t \in Q\} \\
&= \lim_{t_n \downarrow x} f(t_n)
\end{aligned}$$

Then $\tilde{f}$ is increasing and right continuous.

Proof. The function $\tilde{f}$ is clearly increasing. Let $x_0 \in \mathbb{R}$ and fix $\varepsilon > 0$. We shall show that there is an $x > x_0$ such that

$$0 \leq \tilde{f}(x) - \tilde{f}(x_0) < \varepsilon.$$

By the definition, there exists $t_0 \in Q$ such that $t_0 > x_0$ and

$$f(t_0) - \varepsilon < \tilde{f}(x_0) < f(t_0).$$

Hence

$$|f(t_0) - \tilde{f}(x)| < \varepsilon.$$

Thus if $t \in Q$ is such that $x_0 < t < t_0$, we have

$$0 \leq f(t) - \tilde{f}(x_0) \leq f(t_0) - \tilde{f}(x_0) < \varepsilon.$$

That is, for all $x_0 < t < t_0$, we have

$$f(t) < \tilde{f}(x_0) + \varepsilon$$

and therefore if $x_0 < x < t_0$ we see that

$$0 \leq \tilde{f}(x) - \tilde{f}(x_0) < \varepsilon,$$

proving the right continuity of $\tilde{f}$. $\square$

Theorem 1.5 (Helly's Selection Theorem). *Let $\{F_n\}$ be a sequence of of distribution functions. There exists a subsequence $\{F_{n_k}\}$ and a right continuous nondecreasing function function F such that $F_{n_k}(x) \to F(x)$ for all points x of continuity of F.*

Proof. Let $q_1, q_2, \ldots$ be an enumeration of the rational. The sequence $\{F_n(q_1)\}$ has values in $[0, 1]$. Hence, there exists a subsequence $F_{n_1}(q_1) \to G(q_1)$. Similarly for $F_{n_1}(q_2)$ and so on. schematically we see that

$$
\begin{aligned}
&q_1 \colon F_{n_1}, \ldots \to G(q_1) \\
&q_2 \colon F_{n_2}, \ldots \to G(q_2). \\
&\vdots \\
&q_k \colon F_{n_k}(q_k) \ldots \to G(q_k) \\
&\qquad \vdots
\end{aligned}
$$

Now, let $\{F_{n_n}\}$ be the diagonal subsequence. Let q_j be any rational. Then

$$F_{n_n}(q_j) \to G(q_j).$$

So, we have a nondecreasing function G defined on all the rationals. Set

$$
\begin{aligned}
F(x) &= \inf\{G(q) \colon q \in Q \colon q > x\} \\
&= \lim_{q_n \downarrow x} G(q_n)
\end{aligned}
$$

By the Lemma 1.1 F is right continuous and nondecreasing. Next, let us show that $F_{n_k}(x) \to F(x)$ for all points of continuity of F. Let x be such a point and pick $r_1, r_2, s \in Q$ with $r_1 < r_2 < x < s$ so that

$$
\begin{aligned}
F(x) - \varepsilon < F(r_1) &\le F(r_2) \\
&\le F(x) \le F(s) \\
&< F(x) + \varepsilon.
\end{aligned}
$$

Now, since $F_{n_k}(r_2) \to F(r_2) \geq F(r_1)$ and $F_{n_k}(s) \to F(s)$ we have for n_k large enough,

$$F(x) - \varepsilon < F_{n_k}(r_2) \leq F_{n_k}(x) \leq F_{n_k}(s) < F(x) + \varepsilon$$

and this shows that $F_{n_k}(x) \to F(x)$, as claimed. $\square$

When can we guarantee that the above function is indeed a distribution?

Theorem 1.6. *Every weak subsequential limit μ of $\{\mu_n\}$ is a probability measures if and only if for every $\varepsilon > 0$ there exists a bounded interval $I_\varepsilon = (a, b]$ such that*

$$\inf_n \mu_n(I_\varepsilon) > 1 - \varepsilon. \tag{*}$$

In terms of the distribution functions this is equivalent to the statement that for all $\varepsilon > 0$, there exists an $M_\varepsilon > 0$ such that $\sup_n\{1 - F_n(M_\varepsilon) + F_n(-M_\varepsilon)\} < \varepsilon$. A sequence of probability measures satisfying $(*)$ is said to be tight. Notice that if μ_n is unit mass at n then clearly μ_n is not tight. "The mass of μ_n scapes to infinity." The tightness condition prevents this from happening.

Proof. Let $\mu_{n_k} \Rightarrow \mu$. Let $J \supset I_\varepsilon$ and $\mu(\partial J) = 0$. Then

$$\begin{aligned} \mu(\mathbb{R}) \geq \mu(J) &= \lim_{n\to\infty} \mu_{n_k}(J) \\ &\geq \overline{\lim} \mu_{n_k}(I_\varepsilon) \\ &> 1 - \varepsilon. \end{aligned}$$

Therefore, $\mu(\mathbb{R}) = 1$ and μ is a probability measure.

Conversely, suppose $(*)$ fails. Then we can find an $\varepsilon > 0$ and a sequence n_k such that

$$\mu_{n_k}(I) \leq 1 - \varepsilon,$$

for all n_k and all bounded intervals I. Let $\mu_{n_{k_j}} \to \mu$ weakly. Let J be a continuity interval for μ. Then

$$\begin{aligned}\mu(J) = \lim_{j\to\infty} \mu_{n_{k_j}}(J) &\leq \underline{\lim}\mu_{n_{k_j}}(J)\\ &\leq 1-\varepsilon.\end{aligned}$$

Therefore, $\mu(\mathbb{R}) \leq 1-\varepsilon$ and μ is not a probability measure. $\square$

§2 Characteristic Functions.

Let μ be a probability measure on $\mathbb{R}$ and define its Fourier transform by $\hat{\mu}(t) = \int\limits_{\mathbb{R}} e^{itx} d\mu(x)$. Notice that the Fourier transform is a a complex valued function satisfying $|\hat{\mu}(t)| \leq \mu(\mathbb{R}) = 1$ for all $t \in \mathbb{R}$. If X be a random variable its characteristic function is defined by

$$\varphi_X(t) = E(e^{itX}) = E(\cos(tX)) + iE(\sin(tX)).$$

Notice that if $\mu$is the distribution measure of X then

$$\varphi_X(t) = \int_{\mathbb{R}} e^{itx} d\mu(x) = \hat{\mu}(t).$$

and again $|\varphi_X(t)| \leq 1$. Note that if $X \sim Y$ then $\varphi_X(t) = \varphi_Y(t)$ and if X and Y are independent then

$$\varphi_{X+Y}(t) = E(e^{itX}e^{itY}) = \varphi_X(t)\varphi_Y(t).$$

In particular, if $X_1, X_2, \ldots, X_n$ are are i.i.d., then

$$\varphi_{X_n}(t) = (\varphi_{X_1}(t))^n.$$

Notice also that if $\overline{(a+ib)} = a - ib$ then $\overline{\varphi_X(t)} = \varphi_X(-t)$. The function φ is uniformly continuous. To see the this observe that

$$\begin{aligned}|\varphi(t+h) - \varphi(t)| &= |E|e^{i(t+h)X} - e^{itX})|\\ &\leq E|e^{ihX} - 1|\end{aligned}$$

and use the continuity of the exponential to conclude the uniform continuity of φ_X. Next, suppose a and b are constants. Then

$$\varphi_{aX+b}(t) = e^{itb}\varphi_X(at).$$

In particular,

$$\varphi_{-X}(t) = \varphi_X(-t) = \overline{\varphi_X(t)}.$$

If $-X \sim X$ then $\overline{\varphi_X(t)} = \varphi_X(t)$ andφ_X is real. We now proceed to present some examples which will be useful later.

Examples 2.1.

(i) (Point mass at a) Suppose $X \sim F = \delta_a$. Then

$$\varphi(t) = E(e^{itX}) = e^{ita}$$

(ii) (Coin flips) $P(X = 1) = P(X = -1) = 1/2$. Then

$$\varphi(t) = E(e^{itX}) = \frac{1}{2}e^{it} + \frac{1}{2}e^{-it} = \frac{1}{2}(e^{it} + e^{-it}) = \cos(t).$$

(iii) (Bernoulli) $P(X = 1) = p,\ P(X = 0) = 1 - p$. Then

$$\begin{aligned}\varphi(t) = E(e^{itX}) &= pe^{it} + (1 - p)\\ &= 1 + p(e^{it} - 1)\end{aligned}$$

(iv) (Poisson distribution) $P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!},\ k = 0, 1, 2, 3\ldots$

$$\begin{aligned}\varphi(t) &= \sum_{k=0}^{\infty} e^{itk}\frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda e^{it})^k}{k!}\\ &= e^{-\lambda}e^{\lambda e^{it}}\\ &= e^{\lambda(e^{it}-1)}\end{aligned}$$

(v) (Exponential) Let X be exponential with density e^{-y}. Integration by parts gives

$$\varphi(t) = \frac{1}{(1-it)}.$$

(vi) (Normal) $X \sim N(0,1)$.

$$\varphi(t) = e^{-t^2/2}.$$

Proof of (vi). Writing $e^{itx} = \cos(tx) + i\sin(tx)$ we obtain

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{itx} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \cos txe^{-x^2/2} dx$$

$$\begin{aligned}
\varphi'(t) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} -x\sin(tx)e^{-x^2/2} dx \\
&= -\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \sin(tx)xe^{-x^2/2} dx \\
&= -\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} t\cos(tx)e^{-x^2/2} dx \\
&= -t\varphi(t).
\end{aligned}$$

This gives $\frac{\varphi'(t)}{\varphi(t)} = -t$ which, together with the initial condition $\varphi(0) = 1$, immediately yields $\varphi(t) = e^{-t^2/2}$ as desired. □

Theorem 2.1 (The Fourier Inversion Formula). *Let μ be a probability measure and let $\varphi(t) = \int_{\mathbb{R}} e^{itx} d\mu(x)$. Then if $x_1 < x_2$*

$$\mu(x_1, x_2) + \frac{1}{2}\mu(x_1) + \frac{1}{2}\mu(x_2) = \lim_{T\to 0} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-itx_1} - e^{-itx_2}}{it} \varphi(t) dt.$$

Remark. The existence of the limit is part of the conclusion. Also, we do not mean that the integral converges absolutely. For example, if $\mu = \delta_0$ then $\varphi(t) = 1$. If $x_1 = -1$ and $x_2 = 1$, then we have the integral of $\dfrac{2\sin t}{t}$ which does not converse absolutely.

Recall that

$$\operatorname{sign}(\alpha) = \begin{cases} 1 & \alpha > 0 \\ 0 & \alpha = 0 \\ -1 & \alpha < 0 \end{cases}$$

Lemma 2.1. *For all $y > 0$,*

$$0 \le sign(\alpha) \int_0^y \frac{\sin(\alpha x)}{x}\, dx \le \int_0^\pi \frac{\sin x}{x}\, dx, \tag{2.1}$$

$$\int_0^\infty \frac{\sin(\alpha x)}{x}\, dx = \pi/2\ sign(\alpha), \tag{2.2}$$

$$\int_0^\infty \frac{1-\cos\alpha x}{x^2}\, dx = \frac{\pi}{2}|\alpha|. \tag{2.3}$$

Proof. Let $\alpha x = u$. It suffices to prove (2.1)–(2.3) for $\alpha = 1$. For (2.1), write $[0,\infty) = [0,\pi]\cup[\pi,2\pi],\ldots$ and choose n so that $n\pi < y \le (n+1)\pi$. Then

$$\int_0^y \frac{\sin x}{x}\, dx = \sum_{k=0}^{n}\left(\int_{k\pi}^{(k+1)\pi} \frac{\sin x}{x} dx\right) + \int_{n\pi}^{y} \frac{\sin x}{x}\, dx$$
$$= \int_0^\pi \frac{\sin x}{x}\, dx + \int_\pi^{2\pi} \frac{\sin x}{x}\, dx + \int_{2\pi}^{3\pi} \frac{\sin x}{x} dx + \ldots + \int_{n\pi}^{y} \frac{\sin x}{x} dx$$
$$= \int_0^\pi \frac{\sin x}{x} dx + (-1)a_1 + (-1)^2 a_2 + \ldots + (-1)^{n-1}a_{n-1} + (-1)^n \int_{n\pi}^{y} \frac{\sin x}{x} dx$$

where $|a_{j+1}| < |a_j|$. If n is odd then $n-1$ is even and $\int\limits_{n\pi}^{y} \frac{\sin x}{x} dx < 0$. Comparing terms we are done. If n is even, the result follows by replacing y with $(n+1)\pi$ and using the same argument.

For (2.2) and (2.3) apply Fubini's Theorem to obtain

$$\begin{aligned}\int_0^\infty \frac{\sin x}{x} dx &= \int_0^\infty \sin x \int_0^\infty e^{-ux} du dx \\ &= \int_0^\infty \left(\int_0^\infty e^{-ux} \sin x dx\right) du \\ &= \int_0^\infty \left(\frac{du}{1+u^2}\right) \\ &= \pi/2.\end{aligned}$$

and

$$\begin{aligned}\int_0^\infty \frac{1-\cos x}{x^2}\,dx &= \int_0^\infty \frac{1}{x^2}\int_0^x \sin u du dx \\ &= \int_0^\infty \sin u \int_u^\infty \frac{dx}{x^2}\,du \\ &= \int_0^\infty \frac{\sin u}{u} du \\ &= \pi/2.\end{aligned}$$

This completed the proof. $\square$

Proof of Theorem 2.1. We begin by observing that

$$\left|\frac{e^{it(x-x_1)} - e^{it(x-x_2)}}{it}\right| = \left|\int_{x_1}^{x_2} e^{-itu} du\right| \le |x_1 - x_2|$$

and hence for any $T > 0$,

$$\int_{\mathbb{R}^1}\int_{-T}^{T} |x_2 - x_2| dt d\mu(x) \le 2T|x_1 - x_2| < \infty.$$

From this, the definition of φ and Fubini's Theorem, we obtain

$$\begin{aligned}\frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-i+x_1} - e^{-itx_2}}{it}\,\varphi(t)dt &= \int_{-\infty}^{\infty}\int_{-T}^{T} \frac{e^{-itx_1} - e^{-itx_2}}{2\pi it}\,e^{itx} dt d\mu(x) \\ &= \int_{-\infty}^{\infty}\left[\int_{-T}^{T} \frac{e^{it(x-x_1)} - e^{it(x-x_2)}}{2\pi it} dt\right] d\mu(x) \\ &= \int_{-\infty}^{\infty} F(T, x, x_1, x_2) d\mu(x) \qquad (2.4)\end{aligned}$$

Now,

$$\begin{aligned}F(T, x, x_1, x_2) &= \frac{1}{2\pi i}\int_{-T}^{T} \frac{\cos(t(x-x_1))}{t}\,dt + \frac{i}{2\pi i}\int_{-T}^{T} \frac{\sin(t(x-x_1))}{t}\,dt \\ &- \frac{1}{2\pi i}\int_{-T}^{T} \frac{\cos(t(x-x_2))}{t}\,dt - \frac{i}{2\pi i}\int_{-T}^{T} \frac{\sin(t(x-x_2))}{t}\,dt \\ &= \frac{1}{\pi}\int_0^T \frac{\sin(t(x-x_1))}{t}\,dt - \frac{1}{\pi}\int_0^T \frac{\sin(t(x-x_2))}{t}\,dt,\end{aligned}$$

using the fact that $\dfrac{\sin(t(x-x_i))}{t}$ is even and $\dfrac{\cos(t(x-x_i)}{t}$ odd.

By (2.1) and (2.2),

$$|F(T,x,x_1,x_2)| \le \frac{2}{\pi}\int_0^{\pi} \frac{\sin t}{t} dt.$$

and

$$\lim_{T\to\infty} F(T,x,x_1,x_2) = \begin{cases} -\frac{1}{2}-(-\frac{1}{2})=0, & \text{if } x<x_1 \\ 0-(-\frac{1}{2})=\frac{1}{2}, & \text{if } x=x_1 \\ \frac{1}{2}-(-\frac{1}{2})=1, & \text{if } x_1<x<x_2 \\ \frac{1}{2}-0=\frac{1}{2} & \text{if } x=x_2 \\ \frac{1}{2}-\frac{1}{2}=0, & \text{if } x>x_2 \end{cases}$$

Therefore by the dominated convergence theorem we see that the right hand side of (2.4) is

$$\int_{(-\infty,x_1)} 0\cdot d\mu + \int_{\{x_1\}} \frac{1}{2}\ d\mu + \int_{(x_1,x_2)} 1\cdot d\mu + \int \frac{1}{2}\ d\mu + \int_{(x_2,\infty)} 0\cdot d\mu$$
$$= \mu(x_1,x_2) + \frac{1}{2}\ \mu\{x_1\} + \frac{1}{2}\ \mu\{x_2\},$$

proving the Theorem. □

Corollary 2.1. *If two probability measures have the same characteristic function then they are equal.*

This follows from the following

Lemma 2.1. *Suppose The two probability measures μ_1 and μ_2 agree on all intervals with endpoints in a given dense sets, then they agree on all of $\mathcal{B}(R)$.*

This follows from our construction, (see also Chung, page 28).

Proof Corollary 2.1. Since the atoms of both measures are countable, the two measures agree, the union of their atoms is also countable and hence we may apply the Lemma. □

Corollary 2.2. *Suppose X is a random variable with distribution function F and characteristic function satisfying $\int_{\mathbb{R}} |\varphi_X| dt < \infty$. Then F is continuously differentiable and*

$$F'(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-ity} \varphi_X(y) dy.$$

Proof. Let $x_1 = x - h,\ x_2 = x,\ h > 0$. Since $\mu(x_1, x_2) = F(x_2-) - F(x_1)$ we have

$$\begin{aligned}
&F(x_2-) - F(x_1) + \frac{1}{2}(F(x_1) - F(x_1-)) + \frac{1}{2}(F(x_2) - F(x_2-)) \\
&= \mu(x_1, x_2) + \frac{1}{2}\mu\{x_1\} + \frac{1}{2}\mu\{x_2\} \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\frac{e^{-it(x-h)} - e^{-itx}}{it} \right) \varphi_X(t) dt.
\end{aligned}$$

Since

$$\left| \frac{e^{-it(x-h)} - e^{-itx}}{it} \right| = \left| \int_{x-h}^{x} e^{-ity} dy \right| \le h$$

we see that

$$\begin{aligned}
&\lim_{h\to\infty} (\mu(x_1, x_2) + \frac{1}{2}\mu(x_1)) + \frac{1}{2}\mu\{x_2\} \le \\
&\le \lim_{h\to 0} \frac{h}{2\pi} \int_{\mathbb{R}} |\varphi_X(t)| = 0.
\end{aligned}$$

Hence, $\mu\{x\} = 0$ for any $x \in \mathbb{R}$, proving the continuity of F. Now,

$$\begin{aligned}
\frac{F(x+h) - F(x)}{h} &= \mu(x, x+h) \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} \left(\frac{e^{-it} - e^{-it(x+h)}}{hit} \right) \varphi_X(t) dt \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} -\frac{(e^{-it(x+h)} - e^{-itx})}{hit}\ \varphi_X(t) dt.
\end{aligned}$$

Let $h \to 0$ to arrive at

$$F'(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \varphi_X(t) dt.$$

Note that the continuity of F' follows from this, he continuity of the exponential and the dominated convergence theorem. $\square$

Writing
$$F(x) = \int_{-\infty}^{x} F'(t)dt = \int_{-\infty}^{x} f(t)dt$$
we see that it has a density
$$f = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\varphi_X(t)dt.$$
and hence also
$$\varphi(t) = \int_{\mathbb{R}} e^{itx} f(x)dx.$$

§3 Weak convergence and characteristic functions.

Theorem 3.1. *Let $\{\mu_n\}$ be a sequence of probability measures with characteristic functions φ_n. (i) If μ_n converges weakly to a probability measure μ with characteristic function φ, then $\varphi_n(t) \to \varphi(t)$ for all $t \in \mathbb{R}$. (ii) If $\varphi_n(t) \to \tilde{\varphi}(t)$ for all $t \in \mathbb{R}$ where $\tilde{\varphi}$ is a continuous function at 0, then the sequence of measures $\{\mu_n\}$ is tight and converges weakly to a measure μ and $\tilde{\varphi}$ is the characteristic function of μ. In particular, if $\varphi_n(t)$ converges to a characteristic function φ then $\mu_n \Rightarrow \mu$.*

Example 3.1. Let $\mu_n \sim N(0, n)$. Then $\varphi_n(t) = e^{-\frac{nt^2}{2}}$. (By scaling if $X \sim N(\mu, \sigma^2)$ then $\varphi_X(t) = e^{i\mu t - \sigma^2 t^2/2}$.) Clearly $\varphi_n \to 0$ for all $t \neq 0$ and $\varphi_n(0) = 1$ for all n. Thus $\varphi_n(t)$ converges for ever t but the limit is not continuous at 0. Also with
$$\mu_n(-\infty, x] = \frac{1}{\sqrt{2\pi n}}\int_{-\infty}^{x} e^{-t^2/2n}dt$$
a simple change of variables ($r = \frac{t}{\sqrt{n}}$) gives
$$\mu_n = (-\infty, x] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{x}{\sqrt{n}}} e^{\frac{-t^2}{2}}dt \to 1/2$$
and hence no weak convergence.

Proof of (i). This is the easy part. Note that $g(x) = e^{itx}$ is bounded and continuous. Since $\mu_n \Rightarrow \mu$ we get that $E(g(X_n)) \to E(y(X_\infty))$ and this gives $\varphi_n(t) \to \varphi(t)$ for every $t \in \mathbb{R}$.

For the proof of (ii) we need the following Lemma.

Lemma 3.1 (Estimate of μ in terms of φ). *For all $A > 0$ we have*

$$\mu[-2A, 2A] \geq A\left|\int_{-A^{-1}}^{A^{-1}} \varphi(t)dt\right| - 1. \tag{3.1}$$

This, of course can also be written as

$$1 - A\left|\int_{-A^{-1}}^{A^{-1}} \varphi(t)|dt| \geq -\mu[-2A, 2A], \tag{3,2}$$

or

$$P\{|X| > 2A\} \leq 2 - A\left|\int_{-A^{-1}}^{A^{-1}} \varphi(t)dt\right|. \tag{3.3}$$

Proof of (ii). Let $\delta > 0$.

$$\begin{aligned}\left|\frac{1}{2\delta}\int_{-\delta}^{\delta} \varphi(t)dt\right| \leq \left|\frac{1}{2\delta}\int_{-\delta}^{\delta} \varphi_n(t)dt\right| \\ + \frac{1}{2\delta}\int_{-\delta}^{\delta} |\varphi_n(t) - \varphi(t)|dt.\end{aligned}$$

Since $\varphi_n(t) \to \varphi(t)$ for all t, we have for each fixed $\delta > 0$ (by the dominated convergence theorem)

$$\lim_{n\to\infty} \frac{1}{2\delta}\int\limits_{-\delta}^{\delta} |\varphi_n(t) - \varphi(t)|dt \to 0.$$

Since φ is continuous at 0, $\lim\limits_{\delta\to 0} \dfrac{1}{2\delta}\displaystyle\int_{\delta}^{\delta} |\varphi(t)|dt = |\varphi(0)| = 1$. Thus for all $\varepsilon > 0$ there exists a $\delta = \delta(\varepsilon) > 0$ and $n_0 = n_0(\varepsilon)$ such that for all $n \geq n_0$,

$$1 - \varepsilon/2 < \left|\frac{1}{2\delta}\int_{-\delta}^{\delta} \varphi_n(t)dt\right| + \varepsilon/2,$$

or

$$\frac{1}{\delta}\left|\int_{-\delta}^{\delta}\varphi_n(t)dt\right| > 2(1-\varepsilon).$$

Applying the Lemma with $A = \frac{1}{\delta}$ gives

$$\mu_n[-2\delta^{-1},\ 2\delta^{-1}] > \delta\left|\int_{-\delta}^{\delta}\varphi(t)dt\right| > 2(1-\varepsilon) - 1 = 1 - 2\varepsilon,$$

for all $n \geq n_0$. Thus the sequence $\{\mu_n\}$ is tight. Let $\mu_{n_k} \Rightarrow \nu$. Then ν a probability measure. Let ψ be the characteristic function of ν. Then since $\mu_{n_k} \Rightarrow \nu$ the first part implies that $\varphi_{n_k}(t) \to \psi(t)$ for all t. Therefore, $\psi(t) = \tilde{\varphi}(t)$ and hence $\tilde{\varphi}(t)$ is a characteristic function and any weakly convergent subsequence musty converge to a measure whose characteristic function is $\tilde{\varphi}$. This completes the proof. $\square$

Proof of Lemma 3.1. For any $T > 0$

$$\begin{aligned}\int_{-T}^{T}(1 - e^{itx})dt &= 2T - \int_{-T}^{T}(\cos tx + i\sin tx)dt \\ &= 2T - \frac{2\sin(Tx)}{x}.\end{aligned}$$

Therefore,

$$\frac{1}{T}\int_{\mathbb{R}^n}\int_{-T}^{T}(1 - e^{itx})dtd\mu(x) = 2 - \int_{\mathbb{R}}\frac{2\sin(Tx)}{Tx}\ d\mu(x)$$

or

$$2 - \frac{1}{T}\int_{-T}^{T}\varphi(t)dt = 2 - \int_{\mathbb{R}}\frac{2\sin(\pi x)}{Tx}\ d\mu(x).$$

That is, for all $T > 0$,

$$\frac{1}{2T}\ \int_{-T}^{T}\varphi(t)dt = \int_{\mathbb{R}}\frac{\sin(Tx)}{Tx}\ d\mu(x).$$

Now, for any $|x| > 2A$,

$$\left|\frac{\sin(Tx)}{Tx}\right| \leq \frac{1}{|Tx|} \leq \frac{1}{(2TA)}$$

and also clearly,

$$\left|\frac{\sin Tx}{Tx}\right| < 1, \text{for all } x.$$

Thus for any $A > 0$ and any $T > 0$,

$$\begin{aligned}
|\int_{\mathbb{R}} \frac{\sin(Tx)}{Tx}\ d\mu(x)| &= \left|\int_{-2A}^{2A} \frac{\sin(Tx)}{Tx}\ \mu(dx) + \int_{|x|>2A} \frac{\sin(Tx)}{Tx}\ d\mu(x)\right| \\
&\le \mu[-2A, 2A] + \frac{1}{2TA}[1 - \mu[-2A, 2A]] \\
&= \left[1 - \frac{1}{2TA}\right]\mu[-2A, 2A] + \frac{1}{2TA}.
\end{aligned}$$

Now, take $T = A^{-1}$ to conclude that

$$\begin{aligned}
\frac{A}{2}\left|\int_{-A^{-1}}^{A^{-1}} \varphi(t)dt\right| &\le \left|\int_{\mathbb{R}} \frac{\sin Tx}{Tx}\ d\mu\right| \\
&= \frac{1}{2}\mu[-2A, 2A] + 1/2
\end{aligned}$$

which completes the proof. □

Corollary. $\mu\{x: |x| > 2/T\} \le \frac{1}{T}\ \int_{-T}^{T}(1 - \varphi(t))dt$, *or in terms of the random variable,*

$$P\{|X| > 2/T\} \le \frac{1}{T}\int_{-T}^{T}(1 - \varphi(t))dt,$$

or

$$P\{|X| > T\} \le T/2\int_{-T^{-1}}^{T^{-1}}(1 - \varphi(t))dt.$$

§4 Moments and Characteristic Functions.

Theorem 4.1. *Suppose X is a random variable with $E|X|^n < \infty$ for some positive integer n. Then its characteristic function φ has bounded continuous derivatives of any order less than or equal to n and*

$$\varphi^{(k)}(t) = \int_{-\infty}^{\infty} (ix)^k e^{itx} d\mu(x),$$

for any $k \leq n$.

Proof. Let μ be the distribution measure of X. Suppose $n = 1$. Since $\int_{\mathbb{R}} |x| d\mu(x) < \infty$ the dominated convergence theorem implies that

$$\begin{aligned} \varphi'(t) &= \lim_{h \to 0} \frac{\varphi(t+h) - \varphi(t)}{h} = \lim_{n \to 0} \int_{\mathbb{R}} \left(\frac{e^{i(t+h)x} - e^{itx}}{h} \right) d\mu \\ &= \int_{\mathbb{R}} (ix) e^{itx} d\mu(x). \end{aligned}$$

We now continue by induction to complete the proof. $\square$

Corollary 4.1. *Suppose* $E|X|^n < \infty$, n *an integer. Then its characteristic function* φ *has the following Taylor expansion in a neighborhood of* $t = 0$.

$$\varphi(t) = \sum_{m=0}^{n} i^m \frac{t^m E(X)^m}{m!} + o(t^n).$$

We recall here that $g(t) = o(t^m)$ as $t \to 0$ means $g(t)/t^m \to 0$ as $t \to 0$.

Proof. By calculus, if φ has n continuous derivatives at 0 then

$$\varphi(t) = \sum_{m=0}^{n} \frac{\varphi^{(m)}(0)}{m!} t^m + o(t^n).$$

In the present case, $\varphi^{(m)}(0) = i^m E(X^m)$ by the above theorem.

Theorem 4.2. *For any random variable* X *and any* $n \geq 1$

$$\begin{aligned} \left| Ee^{itX} - \sum_{m=0}^{n} \frac{E(itX)^m}{m!} \right| &\leq E \left| e^{itX} - \sum_{m=0}^{n} \frac{(itX)^m}{m!} \right| \\ &\leq E \left(\min \left(\frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!} \right) \right). \end{aligned}$$

This follows directly from

Lemma 4.2. *For any real x and any $n \geq 1$,*

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

We note that this is just the Taylor expansion for e^{ix} with some information on the error.

Proof. For all $n \geq 0$ (by integration by parts),

$$\int_0^x (x-s)^n e^{is} ds = \frac{x^{n+1}}{(n+1)} + \frac{i}{(n+1)} \int_0^x (x-s)^{n+1} e^{is} ds.$$

For $n = 0$ this is the same as

$$\frac{1}{i}(e^{ix} - 1) = \int_0^x e^{is} ds = x + i \int_0^x (x-s) e^{is} ds$$

or

$$e^{ix} = 1 + ix + i^2 \int_0^x (x-s) e^{is} ds.$$

For $n = 1$,

$$e^{ix} = 1 + ix + \frac{i^2 x^2}{2} + \frac{i^3}{2} \int_0^x (x-s)^2 e^{is} ds$$

and continuing we get for any n,

$$e^{ix} - \sum_{m=0}^{n} \frac{(ix)^m}{m!} = \frac{i^{n+1}}{n!} \int_0^x (x-s)^n e^{is} ds.$$

So, need to estimate the right hand side.

$$\left| \frac{i^{n+1}}{n!} \int_0^x (x-s)^n e^{is} ds \right| \leq \frac{1}{n!} \left| \int_0^x (x-s)^n dx \right| = \frac{|x|^{n+1}}{(n+1)!}.$$

This is good for $|x|$ small. Next,

$$\frac{i}{n} \int_0^x (x-s)^n e^{is} ds = -\frac{x^n}{n} + \int_0^x (x-s)^{n-1} e^{is} ds.$$

Since

$$\frac{x^n}{n} = \int_0^x (x-s)^{n-1} ds$$

we set

$$\frac{i}{n}\int_0^x (x-s)^n e^{is} ds = \int_0^x (x-s)^{n-1}(e^{is}-1) ds$$

or

$$\frac{i^{n+1}}{n!}\int_0^x (x-s)^n e^{is} ds = \frac{i^n}{(n-1)!}\int_0^x (x-s)^{n-1}(e^{is}-1) ds.$$

This gives

$$\begin{aligned}\left|\frac{i^{n+1}}{n!}\int_0^x (x-s)^n e^{is} ds\right| &\le \frac{2}{(n-1)!}\int_0^{|x|} (x-s)^{n-1} ds \\ &\le \frac{2}{n!}|x|^n,\end{aligned}$$

and this completes the proof. □

Corollary 1. *If $EX = \mu$ and $E|X|^2 = \sigma^2 < \infty$, then $\varphi(t) = 1 + it\mu - \frac{t^2\sigma^2}{2} + o(t)^2$, as $t \to 0$.*

Proof. Applying Theorem 4.1 with $n = 2$ gives

$$\left|\varphi(t) - \left(1 + it\mu - \frac{t^2\sigma^2}{2}\right)\right| \le t^2 E\left(\frac{|t||X|^3}{3!} \wedge \frac{2|X|^2}{2!}\right)$$

and the expectation goes to zero as $t \to 0$ by the dominated convergence theorem. □

§5. The Central Limit Theorem.

We shall first look at the i.i.d. case.

Theorem 5.1. *$\{X_i\}$ i.i.d. with $EX_i = \mu$, $var(X_i) = \sigma^2 < \infty$. Set $S_n = X_1 + \ldots + X_n$. Then*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \Rightarrow N(0,1).$$

Equivalently, for any real number x,

$$P\{\frac{S_n - \mu}{\sigma\sqrt{n}} \leq x\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dt.$$

Proof. By looking at $X_i' = X_i - \mu$, we may assume $\mu = 0$. By above

$$\varphi_{X_1}(t) = 1 - \frac{t^2\sigma^2}{2} + g(t)$$

with $\frac{g(t)}{t^2} \to 0$ as $t \to 0$. By i.i.d.,

$$\varphi_{S_n}(t) = \left(1 - \frac{t^2\sigma^2}{2} + g(t)\right)^n$$

or

$$\varphi_{\frac{S_n}{\sigma\sqrt{n}}}(t) = \varphi_{S_n}(\sigma\sqrt{n}t) = \left(1 - \frac{t^2}{2n} + g\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n.$$

Since $\dfrac{g(t)}{t^2} \to 0$ as $t \to 0$, we have (for fixed t) that

$$\frac{g\left(\frac{t}{\sigma\sqrt{n}}\right)}{(1/\sqrt{n})^2} = \frac{g\left(\frac{t}{\sigma\sqrt{n}}\right)}{\frac{1}{n}} \to 0,$$

as $n \to \infty$. This can be written as

$$ng\left(\frac{t}{\sigma\sqrt{n}}\right) \to 0 \text{ as } n \to \infty.$$

Next, set $C_n = -\dfrac{t^2}{2} + ng\left(\dfrac{t}{\sigma\sqrt{n}}\right)$ and $C = -t^2/2$. Apply Lemma 5.1 bellow to get

$$\varphi_{\frac{S_n}{\sigma\sqrt{n}}}(t) = \left(1 - \frac{t^2}{2n} + g(t/\sigma\sqrt{n})\right)^n \to e^{-t^2/2}$$

and complete the proof. $\square$

Lemma 5.1. *If C_n are complex numbers with $C_n \to C \in \mathbb{C}$. Then* $\left(1+\frac{C_n}{n}\right)^n \to e^C$.

Proof. First we claim that if $z_1, z_2, \dots, z_n$ and $\omega_1, \dots, \omega_n$ are complex numbers with $|z_j|$ and $|\omega_j| \le \eta$ for all j, then

$$\left|\prod_{m=1}^{n} z_m - \prod_{m=1}^{n} \omega_m\right| \le \eta^{n-1} \sum_{m=1}^{n} |z_m - \omega_m|. \tag{5.1}$$

If $n=1$ the result is clearly true for $n=1$; with equality, in fact. Assume it for $n-1$ to get

$$\begin{aligned}
&\left|\prod_{m=1}^{n} z_m - \prod_{m=1}^{n} \omega_m\right| \le \left|z_n \prod_{m=2}^{n-1} z_m - \omega_n \prod_{m=1}^{n-1} \omega_m\right| \\
&\left|z_n \prod_{m=1}^{n-1} z_m - z_n \prod_{m=1}^{n-1} \omega_m + z_n \prod_{m=1}^{n-1} \omega_m - \omega_n \prod_{m=1}^{n-1} \omega_m\right| \\
&\le \eta\left|\prod_{m=1}^{n-1} z_m - \prod_{m=1}^{n-1} \omega_m\right| + \left|\prod_{m=1}^{n-1} \omega_m\right| |z_n - \omega_m| \\
&\le \eta\eta^{n-2} \sum_{m=1}^{n-1} |z_m - \omega_m| + \eta^{n-1}|z_n - \omega_m| \\
&= \eta^{n-1} \sum_{m=1}^{n} |z_m - \omega_m|.
\end{aligned}$$

Next, if $b \in \mathbb{C}$ and $|b| \le 1$ then

$$|e^b - (1+b)| \le |b|^2. \tag{5.2}$$

For this, write $e^b = 1 + b + \frac{b^2}{2} + \frac{b^3}{3!} + \dots$. Then

$$\begin{aligned}
|e^b - (1+b)| &\le \frac{|b|^2}{2}\left(1 + \frac{2|b|}{3!} + \frac{2|b|^2}{4!} + \dots\right) \\
&\le \frac{|b|^2}{2}\left(1 + \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \dots\right) = |b|^2,
\end{aligned}$$

which establishes (5.2).

With both (5.1) and (5.2) established we let $\varepsilon > 0$ and choose $\gamma > |C|$. Take n large enough so that $|C_n| < \gamma$ and $\gamma^2 e^{2\gamma}/n < \varepsilon$ and $\left|\dfrac{C_n}{n}\right| \leq 1$. Set $z_i = (1 + \frac{C_n}{n})$ and $\omega_i = (e^{C_n/n})$ for all $i = 1, 2, \ldots, n$. Then

$$|z_i| = \left|1 + \frac{C_n}{n}\right| \leq \left(1 + \frac{\gamma}{n}\right) \text{ and } |\omega_i| \leq e^{\gamma/n}$$

hence for both z_i and $|\omega_i|$ we have the bound $e^{\gamma/n}\left(1 + \frac{\gamma}{n}\right)$. By (5.1) we have

$$\begin{aligned} \left|\left(1 + \frac{C_n}{n}\right)^n - e^{C_n}\right| &\leq e^{\frac{\gamma}{n}(n-1)}\left(1 + \frac{\gamma}{n}\right)^{n-1} \sum_{m=1}^{n} \left|e^{\frac{C_n}{n}} - \left(1 + \frac{C_n}{n}\right)\right| \\ &\leq e^{\frac{\gamma}{n}(n-1)}\left(1 + \frac{\gamma}{n}\right)^{n-1} n\left|e^{\frac{C_n}{n}} - \left(1 + \frac{C_n}{n}\right)\right| \end{aligned}$$

Setting $b = C_n/n$ and using (5.2) we see that this quantity is dominated by

$$\begin{aligned} &\leq e^{\frac{\gamma}{n}(n-1)}\left(1 + \frac{\gamma}{n}\right)^{n-1} n\left|\frac{C_n}{n}\right|^2 \\ &\leq \frac{\gamma^2 e^{\frac{\gamma}{n}(n-1)}\left(1 + \frac{\gamma}{n}\right)^{n-1}}{n} \\ &\leq \frac{\gamma^2 e^{\frac{\gamma}{n}(n-1)} e^{\gamma}}{n} \leq \frac{\gamma^2 e^{2\gamma}}{n} < \varepsilon, \end{aligned}$$

which proves the lemma $\square$

Example 5.1. Let X_i be i.i.d. Bernoullians 1 and 0 with probability 1/2. Let $S_n = X_1 + \ldots + X_n$ = total number of heads after n–tones.

$$EX_i = 1/2, \quad \text{var}(X_i) = EX_i^2 - (E(X))^2 = 1/2 - (\frac{1}{4}) = 1/4$$

and hence

$$\frac{S_n - \mu_n}{\sigma\sqrt{n}} = \frac{S_n - \frac{n}{2}}{\sqrt{n/4}} \Rightarrow \chi = N(0,1).$$

From a table of the normal distribution we find that

$$P(\chi > 2) \approx 1 - .9773 = 0.227.$$

Symmetry:

$$P(|\chi| < 2) = 1 - 2(0.227) = .9546.$$

Hence for n large we should have

$$\begin{aligned}
0.95 &\approx P\left(\left|\frac{S_n - \frac{n}{2}}{\sqrt{n^2/4}}\right| < 2\right)\\
&= P\left\{\frac{-2}{2}\sqrt{n} \le S_n - \frac{n}{2} < \frac{2}{2}\sqrt{n}\right\}\\
&= P\left\{\frac{n}{2} - \sqrt{n} < S_n \le \sqrt{n} + n/2\right\}.
\end{aligned}$$

If $n = 250,000$,

$$\begin{aligned}
\frac{n}{2} - \sqrt{n} &= 125,000 - 500\\
\frac{n}{2} + \sqrt{n} &= 125,000 + 5000.
\end{aligned}$$

That is, with probability 0.95, after 250,000 tosses you will get between 124,500 and 125,500 heads.

Examples 5.2. A Roulette wheel has slots 1–38 (18 red and 18 black) and two slots 0 and 00 that are painted green. Players can bet \$1 on each of the red and black slots. The player wins \$1 if the ball falls on his/her slot. Let $X_1, \dots, X_n$ be i.i.d. with $X_i = \{\pm 1\}$ and $P(X_i = 1) = \frac{18}{38}$, $P(X_i = -1) = \frac{20}{38}$. $S_n = X_1 + \cdots + X_n$ is the total fortune of the player after n games. Suppose we want to know $P(S_n \ge 0)$ after large numbers tries. Since

$$\begin{aligned}
E(X_i) &= \frac{18}{38} - \frac{20}{38} = \frac{-2}{38} = -\frac{1}{19}\\
\operatorname{var}(X_i) &= EX^2 - (E(x))^2 = 1 - \left(\frac{1}{19}\right)^2 = 0.9972
\end{aligned}$$

we have

$$P(S_n \ge 0) = P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \ge \frac{-n\mu}{\sigma\sqrt{n}}\right).$$

Take n such that

$$\frac{-\sqrt{n}\left(\frac{-1}{19}\right)}{(.9972)} = 2.$$

This gives $\sqrt{n} = 2(19)(0.9972)$ or $n \approx 3\ 61.4 = 1444$. Hence

$$\begin{aligned} P(S_{1444} \geq 0) &= P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \geq 2\right) \\ &\approx P(\chi \geq 2) \\ &= 1 - 0.9772 \\ &= 0.0228 \end{aligned}$$

Also,

$$\begin{aligned} E(S_{1444}) &= -\frac{1444}{19} = -4.19 \\ &= -76. \end{aligned}$$

Thus, after $n = 1444$ the Casino would have won \$76 of your hard earned dollars, in the average, but there is a probability .0225 that you will be ahead. So, you decide if you want to play!

Lemma 5.2. *Let $C_{n,m}$ be nonnegative numbers with the property that $\max\limits_{1\leq m\leq n} C_{n,m} \to 0$ and $\sum\limits_{m=1}^{n} C_{n,m} \to \lambda$. Then*

$$\prod_{m=1}^{n} (1 - C_{n,m}) \to e^{-\lambda}.$$

Proof. Recall that

$$\lim_{a\downarrow 0} \log \frac{\left(\frac{1}{1-a}\right)}{a} = 1.$$

Therefore, given $\varepsilon > 0$ there exists $\delta > 0$ such that $0 < a < \delta$ implies

$$(1-\varepsilon)a \leq \log\left(\frac{1}{1-a}\right) \leq (1+\varepsilon)a.$$

If n is large enough, $C_{m,n} \le \delta$ and

$$(1-\varepsilon)C_{m,n} \le \log\left(\frac{1}{1-C_{m,n}}\right) \le (1+\varepsilon)C_{m,n}.$$

Thus

$$\sum_{m=1}^{n} \log\left(\frac{1}{1-C_{m,n}}\right) \to \lambda$$

and this is the same as

$$\sum_{m=1}^{n} \log(1-C_{m,n}) \to -\lambda$$

or

$$\log\left(\prod_{m=1}^{n}(1-C_{m,n})\right) \to -\lambda.$$

This implies the result. $\square$

Theorem 5.2 (The Lindeberg–Feller Theorem). *For each n, let $X_{n,m}$, $1 \le m \le n$, be independent r.v.'s with $EX_{n,m} = 0$. Suppose*

(i) $\displaystyle\sum_{m=1}^{n} EX_{n,m}^2 \to \sigma^2,\ \sigma \in (0,\infty).$

(ii) For all $\varepsilon > 0$, $\displaystyle\lim_{n\to\infty}\sum_{m=1}^{n} E(|X_{n,m}|^2;\ |X_{n,m}| > \varepsilon) = 0.$

Then, $S_n = X_{n,1} + X_{n,2} + \ldots + X_{n,m} \Rightarrow N(0,\sigma^2)$.

Example 5.3. Let $Y_1, Y_2, \ldots$ be i.i.d., $EY_i = 0$, $E(Y_i^2) = \sigma^2$. Let $X_{n,m} = Y_m/n^{1/2}$. Then $X_{n,1} + X_{n,2} + \ldots + X_{n,m} = \frac{S_n}{\sqrt{n}}$. Clearly,

$$\sum_{m=1}^{n} \frac{E(Y_m^2)}{n} = \frac{\sigma^2}{n}\sum_{m=1}^{n} 1 = \sigma^2.$$

Also, for all $\varepsilon > 0$,

$$\sum_{m=1}^{n} E(|X_{n,m}|^2;\ |X_{n,m}| > \varepsilon) = nE\left(\frac{|Y_1|^2}{n};\ \frac{|Y_1|}{n^{1/2}} > \varepsilon\right)$$
$$= E(|Y_1|^2;\ |Y_1| > \varepsilon n^{1/2})$$

and this goes to 0 as $n \to \infty$ since $E|Y_1|^2 < \infty$.

Proof. Let $\varphi_{n,m}(t) = E(e^{itX_{n,m}}),\ \sigma^2_{n,m} = E(X^2_{n,m})$. It is enough to show that

$$\prod_{m=1}^{n} \varphi_{n,m}(t) \to e^{-t^2\sigma^2/2}.$$

Let $\varepsilon > 0$ and set $z_{n,m} = \varphi_{n,m}(t),\ \omega_{n,m} = (1 - t^2\sigma^2_{n,m}/2)$. We have

$$\begin{aligned}
|z_{n,m} - \omega_{n,m}| &\le E\left(\frac{|tX_{n,m}|^3}{3!} \wedge \frac{2|tX_{n,m}|^2}{2!}\right) \\
&\le E\left[\frac{|tX_{n,m}|}{3!}^3 \wedge \frac{2|tX_{n,m}|^2}{2!};\ |X_{n,m}| \le \varepsilon\right) \\
&+ E\left(\frac{|tX_{n,m}|^3}{3!} \wedge \frac{2t^2|X_{n,m}|^2}{2!};\ |X_{n,m}| > \varepsilon\right) \\
&\le E\left(\frac{|tX_{n,m}|^3}{3!};\ |X_{n,m}| \le \varepsilon\right) \\
&+ E\left(|tX_{n,m}|^2;\ |X_{n,m}| > \varepsilon\right) \\
&\le \frac{\varepsilon t^3}{6} E|X_{n,m}|^2 + t^2 E(|X_{n,m}|^2;\ |X_{n,m}| > \varepsilon)
\end{aligned}$$

Summing from 1 to n and letting $n \to \infty$ gives (using (i) and (ii))

$$\lim_{n\to\infty} \sum_{m=1}^{n} |z_{n,m} - \omega_{n,m}| \le \frac{\varepsilon t^3 \sigma^2}{6}.$$

Let $\varepsilon \to 0$ to conclude that

$$\lim_{n\to\infty} \sum_{m=1}^{n} |z_{n,m} - \omega_{n,m}| \to 0.$$

Hence with $\eta = 1$ (5.1) gives

$$\left| \prod_{m=1}^{n} \varphi_{n,m}(t) - \prod_{m=1}^{n} \left(1 - \frac{t^2\sigma^2_{n,m}}{2}\right) \right| \to 0,$$

as $n \to \infty$. Now,

$$\sigma^2_{n,m} \le \varepsilon^2 + E(|X_{n,m}|^2;\ |X_{n,m}| > \varepsilon)$$

and therefore,

$$\max_{1\le m\le n}\sigma_{n,m}^2 \le \varepsilon^2 + \sum_{m=1}^{n} E(|X_{n,m}|^2;\ |X_{n,m}|>\varepsilon).$$

The second term goes to 0 as $n\to\infty$. That is, $\max_{1\le m\le n}\sigma_{n,m}^2 \to 0$. Set $C_{n,m} = \frac{t^2\sigma_{n,m}^2}{2}$. Then

$$\sum_{m=1}^{n} C_{n.m} \to \frac{t^2}{2}\sigma$$

and Lemma 5.2 shows that

$$\prod_{m=1}^{n}\left(1-\frac{t^2\sigma_{n,m}}{2}\right) \to e^{-\frac{t^2\sigma^2}{2}},$$

completing the proof of the Theorem. $\square$

We shall now return to the Kolmogorov three series theorem and prove the necessity of the condition. This was not done when we first stated the result earlier. For the sake of completeness we state it in full again.

The Kolmogorov's Three Series Theorem. *Let $X_1, X_2, \dots$ be independent random variables. Let $A>0$ and $Y_m = X_m 1_{(|X_m|\le A)}$. Then $\sum_{n=1}^{\infty} X_n$ converges a.s. if and only if the following three hold:*

(i) $\sum_{n=1}^{\infty} P(|X_n|>A) < \infty$,

(ii) $\sum_{n=1}^{\infty} EY_n$ *converges and*

(iii) $\sum_{n=1}^{\infty}$ *var*$(Y_n) < \infty$.

Proof. We have shown that if (i), (ii), (iii) are true then $\sum_{n=1}^{\infty} X_n$ converges a.s. We now show that if $\sum_{n=1}^{\infty} X_n$ converges then (i)–(iii) hold. We begin by proving (i).

Suppose this is false. That is, suppose

$$\sum_{m=1}^{\infty} P(|X_n| > A) = \infty.$$

Then the Borel–Cantelli lemma implies that

$$P(|X_n| > A \text{ i.o.}) > 0.$$

Thus, $\lim \sum_{m=1}^{n} X_m$ cannot exist. Hence if the series converges we must have (i). Next, suppose (i) holds but $\sum_{n=1}^{\infty} \operatorname{var}(Y_n) = \infty$. Let

$$C_n = \sum_{m=1}^{n} \operatorname{var}(Y_m) \text{ and } X_{n,m} = \frac{(Y_m - EY_m}{C_n^{1/2}}.$$

Then

$$EX_{n,m} = 0 \text{ and } \sum_{m=1}^{n} EX_{n,m}^2 = 1.$$

Let $\varepsilon > 0$ and choose n so large that $\frac{2A}{C_n^{1/2}} < \varepsilon$. Then

$$\begin{aligned} \sum_{m=1}^{n} E(|X_{n,m}|^2;\ |X_{n,m}| > \varepsilon) &\leq \sum_{m=1}^{n} E\left(|X_{n,m}|^2;\ |X_{n,m}| > \frac{2A}{C_n^{1/2}}\right) \\ &\leq \sum_{m=1}^{n} E\left(|X_{n,m}|^2;\ \frac{2A}{C_n^{1/2}} < \frac{|Y_n| + E|Y_m|}{C_n^{1/2}}\right). \end{aligned}$$

But

$$\frac{|Y_n| + E|Y_m|}{C_n^{1/2}} \leq \frac{2A}{C_n^{1/2}}.$$

So, the above sum is zero. Let

$$S_n = X_{n,1} + X_{n,2} + \dots X_{n,m} = \frac{1}{C_n^{1/2}} \sum_{m=1}^{n} (Y_m - EY_m).$$

By Theorem 5.2,

$$S_n \Rightarrow N(0,1).$$

Now, if $\lim_{n\to\infty} \sum_{m=1}^{n} X_m$ exists then $\lim_{n\to\infty} \sum_{m=1}^{n} Y_m$ exists also. (This follows from (i).) Let

$$\begin{aligned} T_n &= \sum_{m=1}^{n} \frac{Y_m}{C_n^{1/2}} \\ &= \frac{1}{C_n^{1/2}} \sum_{m=1}^{n} Y_m \end{aligned}$$

and observe that $T_n \Rightarrow 0$. Therefore, $(S_n - T_n) \Rightarrow \chi$ where $\chi \sim N(0,1)$. (This follows from the fact that $\lim_{n\to\infty} E(g(S_n - T_n)) = \lim_{n\to\infty} E(g(S_n)) = E(g(\chi))$.) But

$$S_n - T_n = -\frac{1}{C_n^{1/2}} \sum_{m=1}^{n} E(Y_m)$$

which is nonrandom. This gives a contradiction and shows that (i) and (iii) hold.

Now, $\sum_{n=1}^{\infty} \operatorname{var}(Y_n) < \infty$ implies $\sum_{m=1}^{\infty} (Y_m - EY_m)$ converges, by the corollary to Kolmogorov maximal inequality. Thus if $\sum_{m=1}^{n} X_n$ converges so does $\sum Y_m$ and hence also $\sum EY_m$. $\square$

§6. The Polya distribution.

We begin with some discussion on the Polya distribution. Consider the density function given by

$$\begin{aligned} f(x) &= (1 - |x|)1_{x\in(-1,1)} \\ &= (1 - |x|)^+. \end{aligned}$$

Its characteristic function is given by

$$\varphi(t) = \frac{2(1 - \cos t)}{t^2}$$

and therefore for all $y \in \mathbb{R}$,

$$\begin{aligned} (1 - |y|)^+ &= \frac{2}{2\pi} \int_{\mathbb{R}} e^{-ity} \frac{(1 - \cos t)}{t^2}\, dt \\ &= \frac{1}{\pi} \int_{\mathbb{R}} \left(\frac{1 - \cos t}{t^2} \right) e^{-ity} dt. \end{aligned}$$

Take $y = -y$ this gives

$$(1 - |y|)^+ = \frac{1}{\pi} \int_{\mathbb{R}} \frac{(1 - \cos t)}{t^2} \, e^{ity} dt.$$

So, if $f_1(x) = \dfrac{1 - \cos x}{\pi x^2}$ which has $\displaystyle\int_{\mathbb{R}} f_1(x) dx = 1$, and we take $X \sim F$ where F has density f_1 we see that $(1 - |t|)^+$ is its characteristic function This is called the Polya distribution. More generally, If $f_a(x) = \frac{1 - \cos ax}{\pi a x^2}$, then then we get the characteristic function $\varphi_a(t) = \left(1 - \left|\frac{t}{a}\right|\right)^+$, just by changing variables. The following fact will be useful below. If $F_1, \ldots, F_n$ have characteristic functions $\varphi_1, \ldots, \varphi_n$, respectively, and $\lambda_i \geq 0$ with $\sum \lambda_i = 1$. Then the characteristic function of $\displaystyle\sum_{i=1}^{n} \lambda_i F_i$ is $\displaystyle\sum_{i=1}^{n} \lambda_i \varphi_i$.

Theorem 6.1 (The Polya Criterion). *Let $\varphi(t)$ be a real and nonnegative function with $\varphi(0) = 1$, $\varphi(t) = \varphi(-t)$, decreasing and convex on $(0, \infty)$ with $\displaystyle\lim_{t \downarrow 0} \varphi(t) = 1$, $\displaystyle\lim_{t \uparrow \infty} \varphi(t) = 0$. There is a probability measure ν on $(0, \infty)$ so that*

$$\varphi(t) = \int_0^\infty \left(1 - \left|\frac{t}{s}\right|\right)^+ d\nu(s)$$

and $\varphi(t)$ is a characteristic function.

Example 6.1. $\varphi(t) = e^{-|t|^\alpha}$ for any $0 < \alpha \leq 2$. If $\alpha = 2$, we have the normal density. If $\alpha = 1$, we have the Cauchy density. Let us in show here that $\exp(-|t|^\alpha)$ is a characteristic function for any $0 < \alpha < 1$. With a more delicate argument, one can do the case $1 < \alpha < 2$. We only need to verify that the function is convex. Differentiating twice this reduces to proving that

$$\alpha t^{2\alpha - 2} - \alpha^2 t^{\alpha - 2} + \alpha t^{\alpha - 2} > 0.$$

This is true if $\alpha^2 t^\alpha - \alpha^2 + \alpha > 0$ which is the same as $\alpha^2 t^\alpha - \alpha^2 + \alpha > 0$ which follows from $\alpha t^\alpha + \alpha(1 - \alpha) > 0$ since $0 < \alpha \leq 1$.

§7. Rates of Convergence; Berry–Esseen Estimates.

Theorem 7.1. *Let X_i be i.i.d., $E|X_i|^2 = \sigma^2$, $EX_i = 0$ and $E|X_i|^3 = \rho < \infty$. If F_n is the distribution of $\dfrac{S_n}{\sigma\sqrt{n}}$ and $\Phi(x)$ is the normal distribution, we have*

$$\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| \le \frac{c\rho}{\sigma^3\sqrt{n}},$$

where c is an absolute constant. In fact, we may take $c = 3$.

More is actually true:

$$F_n(x) = \phi(x) + \frac{H_1(x)}{\sqrt{n}} + \frac{H_2(x)}{n} + \ldots + \frac{H_3(x)}{n^{3/2}} + \ldots$$

where $H_i(x)$ are explicit functions involving Hermit polynomials. We shall not prove this, however.

Lemma 7.1. *Let F be a distribution function and G a real–valued function with the following conditions:*

(i) $\lim_{x\to-\infty} G(x) = 0$, $\lim_{x\to+\infty} G(x) = 1$,

(ii) G has bounded derivative with $\sup_{x\in\mathbb{R}} |G'(x)| \le M$. *Set* $A = \dfrac{1}{2M} \sup_{x\in\mathbb{R}} |F(x) - G(x)|$.

There is a number α such that for all $T > 0$,

$$\begin{aligned} &2MTA\left\{3\int_0^{TA} \frac{1-\cos x}{x^2}\; dx - \pi\right\} \\ &\le \left|\int_{-\infty}^{\infty} \frac{1-\cos Tx}{x^2}\{F(x+\alpha) - G(x+\alpha)\}dx\right|. \end{aligned}$$

Proof. Observe that $A < \infty$, since G is bounded and we may obviously assume that it is positive. Since $F(t) - G(t) \to 0$ at $t \to \pm\infty$, there is a sequence $x_n \to b \in \mathbb{R}$ such that

$$F(x_n) - G(x_n) \to \begin{cases} 2MA \\ or \\ -2MA \end{cases}.$$

Since $F(b) \geq F(b-)$ it follows that either

$$\begin{cases} F(b) - G(b) = 2MA \\ or \\ F(b-) - G(b) = -2MA. \end{cases}$$

Assume $F(b-) - G(b) = -2MA$, the other case being similar.

Put

$$\alpha = b - A < b, \text{ since}$$
$$A = (b - \alpha).$$

If $|x| < A$ we have

$$\begin{aligned} G(b) - G(x + \alpha) &= G'(\xi)(b - \alpha - x) \\ &= G'(\xi)(A - x) \end{aligned}$$

Since $|G'(\xi)| \leq M$ we get

$$\begin{aligned} G(x + a) &= G(b) + (x - A)G'(\xi) \\ &\geq G(b) + (x - A)M. \end{aligned}$$

So that

$$\begin{aligned} F(x + a) - G(x + a) &\leq F(b-) - [G(b) + (x - \Delta)M] \\ &= -2MA - xM + AM \\ &= -M(x + A) \end{aligned}$$

for all $x \in [-A, A]$. Therefore for all $T > 0$,

$$\begin{aligned} \int_{-A}^{A} \frac{1 - \cos Tx}{x^2} \{F(x + \alpha) - G(x + \alpha)\} dx &\leq -M \int_{-A}^{A} \frac{1 - \cos Tx}{x^2} (x + A) dx \\ &= -2MA \int_{0}^{A} \left(\frac{1 - \cos Tx}{x^2} \right) dx \end{aligned}$$

Also,

$$\left|\left\{\int_{-\infty}^{-A}+\int_{A}^{\infty}\right\}\frac{1-\cos Tx}{x^2}\{F(x+\alpha)-G(x+\alpha)\}dx\right| \le 2MA\left\{\int_{-\infty}^{-A}+\int_{A}^{\infty}\right\}\frac{1+\cos Tx}{x^2}dx$$

$$= 4MA\int_{A}^{\infty}\frac{1-\cos Tx}{x^2}\,dx.$$

Adding these two estimates gives

$$\begin{aligned}
&\int_{-\infty}^{\infty}\left(\frac{1-\cos Tx}{x^2}\right)\{F(x+\alpha)-G(x+\alpha)\}dx\\
&\le 2MA\left\{-\int_0^A+2\int_A^\infty\right\}\left\{\frac{1-\cos Tx}{x^2}\right\}dx\\
&= 2MA\left\{-3\int_0^A+2\int_0^\infty\right\}\left\{\frac{1-\cos Tx}{x^2}\right\}dx\\
&= 2MA\left\{-3\int_0^A\frac{1-\cos Tx}{x^2}dx+2\int_0^\infty\frac{1-\cos Tx}{x^2}\,dx\right\}\\
&= 2MA\left\{-3\int_0^A\frac{1-\cos Tx}{x^2}dx+2\left(\frac{\pi T}{2}\right)\right\}\\
&= 2MTA\left\{-3\int_0^{TA}\frac{1-\cos x}{x^2}dx+\pi\right\}<0,
\end{aligned}$$

proving the result. □

Lemma 7.2. *Suppose in addition that G is of bounded variation in $(-\infty,\infty)$ (for example if G has a density) and that*

$$\int_{-\infty}^{\infty}|F(x)-G(x)|dx<\infty.$$

Let $f(t)$ and $g(t)$ be the characteristic functions of F and G, respectively. Then

$$A\le\frac{1}{2\pi M}\int_{-T}^{T}\frac{|f(t)-g(t)|}{t}dt+\frac{12}{T\pi},$$

for any $T>0$.

Proof. Since F and G are of bounded variation,

$$f(t)-g(t)=-it\int_{-\infty}^{\infty}\{F(x)-G(x)\}e^{itx}dx.$$

Therefore,

$$\begin{aligned}\frac{f(t)-g(t)}{-it}\ e^{-it\alpha} &= \int_{-\infty}^{\infty}(F(x)-G(x))e^{-it\alpha+itx}dx\\ &= \int_{-\infty}^{\infty}F(x+\alpha)-G(x+\alpha)e^{itx}dx.\end{aligned}$$

It follows from our assumptions that the right hand side is uniformly bounded in α. Multiply the left hand side by $(T-|t|)$ and integrating gives

$$\begin{aligned}&\int_{-T}^{T}\left\{\frac{f(t)-g(t)}{-it}\right\}e^{-it\alpha}(T-|t|)dt\\ &\qquad= \int_{-T}^{T}\int_{-\infty}^{\infty}\{F(x+\alpha)-G(x+\alpha)\}e^{itx}(T-|t|)dxdt\\ &\qquad= \int_{-\infty}^{\infty}\{F(x+\alpha)-G(x+\alpha)\}\int_{-T}^{T}e^{itx}(T-|t|)dtdx\\ &\qquad= \int_{-\infty}^{\infty}(F(x+\alpha)-G(x+\alpha))\int_{-T}^{T}e^{itx}(T-|t|)dtdx.\\ &\qquad= I\end{aligned}$$

Writing

$$\frac{1-\cos Tx}{x^2}=\frac{1}{2}\int_{-T}^{T}(T-|t|)e^{itx}dt$$

we see that

$$I=2\int_{-\infty}^{\infty}(F(x+\alpha)-G(x+\alpha))\left\{\frac{1-\cos Tx}{x^2}\right\}dx$$

which gives

$$\begin{aligned}&\left|\int_{-\infty}^{\infty}\left\{F(x+\alpha)-G(x+\alpha)\right\}\left\{\frac{1-\cos Tx}{x^2}\right\}dx\right|\\ &\qquad\le\frac{1}{2}\left|\int_{-T}^{T}\frac{f(t)-g(t)}{-it}\ e^{-ita}(T-|t|)dt\right|\\ &\qquad\le T/2\int_{-T}^{T}\left|\frac{f(t)-g(t)}{t}\right|dt\end{aligned}$$

Therefore by Lemma 7.1,

$$2MA\left\{3\int_{0}^{TA}\frac{1-\cos x}{x^2}\ dx-\pi\right\}\le\frac{1}{2}\int_{-T}^{T}\left|\frac{f(t)-g(t)}{t}\right|\ dt$$

However,

$$\begin{aligned}3\int_0^{TA}\frac{1-\cos x}{x^2}\,dx-\pi&=3\int_0^{\infty}\frac{1-\cos x}{x^2}\,dx-3\int_{TA}^{\infty}\frac{1-\cos x}{x^2}\,dx-\pi\\&=3\left(\frac{\pi}{2}\right)-3\int_{TA}^{\infty}\frac{1-\cos x}{x^2}\,dx-\pi\\&\geq\frac{3\pi}{2}-6\int_{TA}^{\infty}\frac{dx}{x^2}-\pi=\frac{\pi}{2}-\frac{6}{TA}\end{aligned}$$

Hence,

$$\begin{aligned}\int_{-T}^{T}\frac{|f(t)-g(t)|}{t}\,dt&\geq 2\left(2MA\left(\frac{\pi}{2}-\frac{6}{TA}\right)\right)\\&=2M\pi A-\frac{24M}{T}\end{aligned}$$

or equivalently,

$$A\leq\frac{1}{2M}\int_{-T}^{T}\left|\frac{f(t)-g(t)}{t}\right|\,dt+\frac{12}{T\pi},$$

which proves the theorem. $\square$

Proof of Theorem 7.1. Without loss of generality, $\sigma^2=1$. Then $\rho\geq 1$. We will apply the above lemmas with

$$F(x)=F_n(x)=P\left(\frac{S_n}{\sqrt{n}}>x\right)$$

and

$$G(x)=\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-y^2/2}dy.$$

Clearly they satisfy the hypothesis of Lemma 7.1 and in fact we may take $M=2/5$ since

$$\sup_{x\in\mathbb{R}}|\Phi'(x)|=\frac{1}{\sqrt{2\pi}}=.39894<2/5.$$

Also clearly G is of bounded variation. We need to show that

$$\int_{\mathbb{R}}|F_n(x)-\Phi(x)|dx<\infty.$$

To see this last fact, note that Clearly,

$$\int_{-1}^{1} |F_n(x) - \Phi(x)| dx < \infty$$

and we need to verify that

$$\int_{-\infty}^{-1} |F_n(x) - \Phi(x)| dx + \int_{1}^{\infty} |F(x) - \Phi(x)| dx < \infty. \tag{7.1}$$

For $x > 0$, $P(|X| > x\} \leq \frac{1}{\lambda^2} E|X|^2$, by Chebyshev's inequality. Therefore,

$$(1 - F_n(x)) = P\bigg(\frac{S_n}{\sqrt{n}} > x\bigg) \leq \frac{1}{x^2} E\bigg|\frac{S_n}{\sqrt{n}}\bigg|^2 < \frac{1}{x^2}$$

and if N denotes a normal random variable with mean zero and variance 1 we also have

$$(1 - \Phi(x)) = P(N > x) \leq \frac{1}{x^2} E|N|^2 = \frac{1}{x^2}.$$

In particular: for $x > 0$, $\max\left((1 - F_n(x)), (1 - \Phi(x))\right) \leq \frac{1}{x^2}$. If $x < 0$ then

$$F_n(x) = P\bigg(\frac{S_n}{\sqrt{n}} < x\bigg) = P\bigg(-\frac{S_n}{\sqrt{n}} > -x\bigg) \leq \frac{1}{x^2} E\bigg|\frac{S_n}{\sqrt{n}}\bigg|^2 = \frac{1}{x^2}$$

and

$$\Phi(x) = P(N < x) \leq \frac{1}{x^2}.$$

Once again, we have $\max(F_n(x), \Phi(x)) \leq \frac{1}{x^2}$ hence for all $x \neq 0$ we have

$$|F(x) - \Phi(x)| \leq \frac{1}{x^2}.$$

Therefore, (7.1) holds and we have verified the hypothesis of both lemmas. We obtain

$$\begin{aligned}
|F_n(x) - \Phi(x)| &\leq \frac{1}{\pi} \int_{-T}^{T} \frac{|\varphi^n(t/\sqrt{n}) - e^{-t^2/2}|}{|t|} \, dt + \frac{24M}{\pi T} \\
&\leq \frac{1}{\pi} \int_{-T}^{T} \frac{|\varphi^n(t/\sqrt{n}) - e^{-t^2/2}|}{|t|} \, dt + \frac{48}{5\pi T}
\end{aligned}$$

Assume n is large and take $T = \frac{4\sqrt{n}}{3\rho}$. Then

$$\frac{48}{5\pi T} = \frac{48 \cdot 3}{5\pi 4\sqrt{n}}\rho = \frac{12 \cdot 3}{5\pi\sqrt{n}}\ \rho = \boxed{\frac{36\rho}{5\pi\sqrt{n}}}.$$

Next we claim

$$\frac{1}{|t|}|\varphi^n(t/\sqrt{n}) - e^{-t^2/2}| \leq \frac{1}{T}e^{-t^2/4}\left\{\frac{2t^2}{9} + \frac{|t|^3}{18}\right\} \tag{7.2}$$

for $-T \leq t \leq T$, $T = 4\sqrt{n}/3\rho$ and $n \geq 10$. If this were the case then

$$\begin{aligned}
&\pi T|F_n(x) - \Phi(x)| \leq \int_{-T}^{T} e^{-t^2/4}\left\{\frac{2t^2}{9} + \frac{|t|^3}{18}\right\}dt + \frac{48}{5} \\
&= \int_{-T}^{T} e^{-t^2/4}\left\{\frac{2t^2}{9} + \frac{|t|^3}{18}\right\}dt + \frac{48}{5} \\
&\leq \frac{2}{9}\int_{-\infty}^{\infty} e^{-t^2/4}t^2 dt + \frac{1}{18}\int_{-\infty}^{\infty} e^{-t^2/4}|t|^3 dt + 9.6 \\
&= I + II + 9.6.
\end{aligned}$$

Since

$$\frac{2}{9}\int_{-\infty}^{\infty} e^{-t^2/4}t^2 dt = \frac{8}{9}\sqrt{\pi}$$

and

$$\begin{aligned}
\frac{1}{18}\int_{-\infty}^{\infty} |t|^3 e^{-t^2/4} dt &= \frac{1}{18}\left\{2\int_0^{\infty} t^3 e^{-t^2/4} dt\right\} \\
&= \frac{1}{18}\left\{2\int_0^{\infty} t^2 \cdot t e^{-t^2/4} dt\right\} \\
&= \frac{1}{18} + \left\{2\int_0^{\infty} 2t \cdot 2e^{-t^2/4} dt\right\} \\
&= \frac{1}{18}\left\{8\int_0^{\infty} t e^{-t^2/4} dt\right\} \\
&= \left\{-16 e^{-t^2/4}\Big|_0^{\infty}\right\}\frac{1}{18} \\
&= \frac{16}{18} = \frac{8}{9}.
\end{aligned}$$

Therefore,

$$\pi T|F_n(x) - \Phi(x)| \leq \left(\frac{8}{9}\sqrt{\pi} + \frac{8}{9} + 9.6\right).$$

This gives,

$$\begin{aligned}|F_n(x) - \Phi(x)| &\leq \frac{1}{\pi T}\left\{\frac{8}{9}(1+\sqrt{\pi}\right) + 9.6\right\} \\ &= \frac{3\rho}{4\sqrt{n}\ \pi}\left\{\frac{8}{9}(1+\sqrt{\pi}) + 9.6\right\} \\ &< \frac{3\rho}{\sqrt{n}}.\end{aligned}$$

For $n \leq 9$, the result is clear since $1 \leq \rho$. It remains to prove (7.2). Recall that that $\left|\varphi(t) - \sum_{m=0}^{n} \frac{E(itX)^m}{m!}\right| \leq E(\min \frac{|tX|^{n+1}}{(n+1)!}\ \frac{2|tX|^n}{n!}\Big)$. This gives

$$\left|\varphi(t) - 1 + \frac{t^2}{2}\right| \leq \frac{\rho|t|^3}{6}$$

and hence

$$|\varphi(t)| \leq 1 - t^2/2 + \frac{\rho|t|^3}{6},$$

for $t^2 \leq 2$.

With $T = \frac{4\sqrt{n}}{3\rho}$, if $|t| \leq T$ then $\frac{\rho|t|}{\sqrt{n}} \leq (4/3) < 2$ and $t/\sqrt{n} = \dfrac{4}{3\rho} < 2$. Thus

$$\begin{aligned}\left|\varphi\left(\frac{t}{\sqrt{n}}\right)\right| &\leq 1 - \frac{t^2}{2n} + \frac{\rho|t|^3}{6n^{3/2}} \\ &= 1 - \frac{t^2}{2n} + \frac{\rho|t|}{6\sqrt{n}}\frac{|t|^2}{n} \\ &\leq 1 - \frac{t^2}{2n} + \frac{4}{18}\frac{t^2}{n} \\ &= 1 - \frac{5t^2}{18n} \\ &\leq e^{\frac{-5t^2}{18n}},\end{aligned}$$

given that $1 - x \leq e^{-x}$. Now, let $z = \varphi(t/\sqrt{n}),\ w = e^{-t^2/2n}$ and $\gamma = e^{\frac{-5t^2}{18n}}$. Then for $n \geq 10$, $\gamma^{n-1} \leq e^{-t^2/4}$ and the lemma above gives

$$|z^n - w^n| \leq n\gamma^{n-1}|z - w|$$

which implies that

$$\begin{aligned}
|\varphi(t/\sqrt{n}) - e^{-t^2/2}| &\le n e^{\frac{-5t^2}{18n}(n-1)} \left|\varphi(t/\sqrt{n}) - e^{-t^2/2n}\right| \\
&\le n e^{-t^2/4} \left|\varphi(t/\sqrt{n}) - 1 + \frac{t^2}{2n} - e^{-t^2/2n} + 1 - \frac{t^2}{2n}\right| \\
&\le n e^{-t^2/4} \left|\varphi(t/\sqrt{n}) - 1 + \frac{t^2}{2n}\right| + n e^{-t^2/4} \left|1 - \frac{t^2}{2n} - e^{-t^2/2n}\right| \\
&\le n e^{-t^2/4} \frac{\rho |t|^3}{6n^{3/2}} + n e^{-t^2/4} \frac{t^4}{2 \cdot 4n^2},
\end{aligned}$$

using the fact that $|e^{-x} - (1-x)| \le \frac{x^2}{2}$, for $0 < x < 1$. We get

$$\begin{aligned}
\frac{1}{|t|} \left|\varphi(t/\sqrt{n}) - e^{-t^2/2}\right| &\le \frac{\rho t^2 e^{-t^2/4}}{6\sqrt{n}} + \frac{e^{-t^2/4}|t|^3}{8n} \\
&= e^{-t^2/4} \left\{ \frac{\rho t^2}{6\sqrt{n}} + \frac{|t|^3}{8n} \right\} \\
&\le \frac{1}{T} e^{-t^2/4} \left\{ \frac{2t^2}{9} + \frac{|t|^3}{18} \right\},
\end{aligned}$$

using $\rho/\sqrt{n} = \frac{4}{3}T$ and $\dfrac{1}{n} = \dfrac{1}{\sqrt{n}} \dfrac{1}{\sqrt{n}} \le \dfrac{4}{3} T \dfrac{1}{3}$, $\rho > 1$ and $n \ge 10$.This completed the proof of (7.2) and the proof of the theorem. $\square$

Let us now take a look at the following question. Suppose F has density f. Is it true that the density of $\dfrac{S_n}{\sqrt{n}}$ tends to the density of the normal? This is not always true. as shown in Feller, volume 2, page 489. However, it is true if add some other conditions. We state the theorem without proof.

Theorem. *Let X_i be i.i.d., $EX_i = 0$ and $EX_i^2 = 1$. If $\varphi \in L^1$, then $\dfrac{S_n}{\sqrt{n}}$ has a density f_n which converges uniformly to $\dfrac{1}{\sqrt{2\pi}}\, e^{-x^2/2} = \eta(x)$.*

§8. Limit Theorems in $\mathbb{R}^d$.

Recall that $\mathbb{R}^d = \{(x_1, \dots, x_d) : x_i \in \mathbb{R}\}$. For any two vectors $x, y \in \mathbb{R}^d$ we will write $x \le y$ if $x_i \le y_i$ for all $i = 1, \dots, d$ and write $x \to \infty$ if $x_i \to \infty$ for all

i. Let $X = (x_1, \dots, x_d)$ be a random vector and defined its distribution function by $F(x) = P(X \le x)$. F has the following properties:

(i) If $x \le y$ then $F(x) \le F(y)$.

(ii) $\lim_{x \to \infty} F(x) = 1, \quad \lim_{x_i \to -\infty} F(x) = 0.$

(iii) F is right continuous. That is, $\lim_{y \downarrow x} F(x) = F(x)$.

The distribution measure is given by $\mu(A) = P(X \in A)$, for all $A \in \mathcal{B}(\mathbb{R}^d)$. However, unlike the situation of the real line, a function satisfying (i) $\leftrightarrow$ (ii) may not be the distribution function of a random vector. Example: we must have:

$$
\begin{aligned}
&P(X \in (a_1, b_1] \times (a_2, b_2]) = F(b_1, b_2) - F(a_1, b_2) \\
&P(a < X_1 \le b_1, a_2 \le X_2 \le b_2) - F(b_1, a_2) + F(a_1, a_2).
\end{aligned}
$$

Need: measure of each vect. ≥ 0,

Example 8.1.

$$
F(x_1, x_2) = \begin{cases} 1, & x_1, x_1 \ge 1 \\ 2/3, & x_1 \ge 1,\ 0 \le x_2 \le 1 \\ 2/3, & x_2 \ge 1,\ 0 \le x_1 < 1 \\ 0, & \text{else} \end{cases}
$$

If $0 < a_1,\ a_2 < 1 \le b_1,\ b_2 < \infty$, then

$$
\begin{aligned}
F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) &= 1 - 2/3 - 2/3 + 0 \\
&= -1/3.
\end{aligned}
$$

Hence the measure has

$$\mu(0,1) = \mu(1,0) = 2/3, \ \mu(1,1) = -1/3$$

which is a signed measure (not a probability measure).

If F is the distribution function of $(X_1, \dots, X_n)$, then $F_i(x) = P(X_i \leq x)$, $x \in \mathbb{R}$ is called the marginal distributions of F. We also see that

$$F_i(x) = \lim_{m \to \infty} F(m, \dots, m, x_i, \dots, m)$$

As in the real line, F has a density if there is a nonnegative function f with

$$\int\limits_{\mathbb{R}^n} f(y)dy = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(y_1, y_2, \dots, y_n) dy_1 \dots dy_n = 1$$

and

$$F(x_1, x_2, \dots, x_d) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f(y) dy_1 \dots dy_2.$$

Definition 8.1. *If F_n and F are distribution functions in $\mathbb{R}^d$, we say F_n converges weakly to F, and write $F_n \Rightarrow F$, if $\lim\limits_{n \to \infty} F_n(x) = F(x)$ for all points of continuity of F. As before, we also write $X_n \Rightarrow X$, $\mu_n \Rightarrow \mu$.*

As in the real line, recall that $\overline{A}$ in the closure of A and A^o is its interior and $\partial A = \overline{A} - A^o$ is its boundary. The following two results are exactly as in the real case. We leave the proofs to the reader.

Theorem (Skorohod) 8.1. *Suppose $X_n \Rightarrow X$. Then there exists a sequence of random vectors Y_n and a random vector Y with $Y_n \sim X_n$ and $Y \sim X$ such that $Y_n \to Y$ a.e.*

Theorem 8.2. *The following statements are equivalent to $X_n \Rightarrow X$.*

(i) $Ef(X_n) \to E(f(X))$ for all bounded continuous functions f.

(iii) For all closed sets K, $\overline{\lim} P(X_n \in K) \leq P(X \in K)$.

(iv) For all open sets G, $\underline{\lim} P(X_n \in G) \geq P(X \in G)$.

(v) For all Borel sets A with $(P(X \in \partial A) = 0$,

$$\lim_{n\to\infty} P(X_n \in A) = P(X_\infty \in A).$$

(vi) Let $f : \mathbb{R}^d \to \mathbb{R}$ be bounded and measurable. Let D_f be the discontinuity points of f. If $P(X \in D_f) = 0$, then $E(f(X_n)) \to E(f(X_\infty))$.

Proof. $X_n \Rightarrow X_\infty \Rightarrow$ (i) trivial. (i) $\Rightarrow$ (ii) trivial. (i) $\Rightarrow$ (ii). Let $d(x, K) = \inf\{|x - y| \colon y \in K\}$. Set

$$\varphi_j(t) = \begin{cases} 1 & t \leq 0 \\ 1 - jt & 0 \leq t \leq j^{-1} \\ 0 & {}^{-1} \leq t \end{cases}$$

and let $f_j(x) = \varphi_j(\text{dist }(x, K))$. The functions f_j are continuous and bounded by 1 and $f_j(x) \downarrow I_K(x)$, since K is closed. Therefore,

$$\begin{aligned} \limsup_{n\to\infty} \mu_n(K) &\leq \lim_{n\to\infty} E(f_j(X_n)) \\ &= E(f_j(X)) \end{aligned}$$

and this last quantity $\downarrow P(X \in K)$ as $j \uparrow \infty$.

That (iii) $\Rightarrow$ (iv) follows by taking complements. For (v) implies convergence in distribution, assume F is continuous at $x = (x_1, \dots, x_d)$, and set $A = (-\infty, x] = (-\infty, x_1] \times \dots (-\infty, x_d]$. We have $\mu(\partial A) = 0$. So, $F_n(x) = F(x_n \in A) \to P(X_\infty \in A) = F(x)$. $\square$

As in the real case, we say that a sequence of measurers μ_n is tight if given $\varepsilon \geq 0$ exists an $M_\varepsilon > 0$ such that

$$\inf_n \mu_n([-M_\varepsilon, M_\varepsilon]^d) \geq 1 - \varepsilon.$$

We remark here that Theorem 1.6 above holds also in the setting of $\mathbb{R}^d$. The characteristic function of the random vector $X = (X_1, \ldots, X_d)$ is defined as $\varphi(t) = E(e^{it\cdot X})$ where $t \cdot X = t_1X_1 + \ldots + t_dX_d$.

Theorem 8.3 (The inversion formula in $\mathbb{R}^d$). *Let $A = [a_1, b_1] \times \ldots \times [a_d, b_d]$ with $\mu(\partial A) = 0$. Then*

$$\begin{aligned}\mu(A) &= \lim_{T\to\infty} \frac{1}{(2\pi)^d} \int_{-T}^{T} \cdots \int_{-T}^{T} \psi_1(t_1)\varphi(t) \ldots \psi_d(t_d)\varphi(t) dt_1, \ldots, dt_2 \\ &= \lim_{T\to\infty} \frac{1}{(2\pi)^d} \int_{[-T,T]^d} \prod_{j=1}^{d} \psi_j(t_j)\varphi(t)dt,\end{aligned}$$

where

$$\psi_j(s) = \left(\frac{e^{isa_j} - e^{-sb_j}}{is}\right),$$

for $s \in \mathbb{R}$.

Proof. Applying Fubini's Theorem we have

$$\begin{aligned}&\int_{[-T,T]^d} \prod_{j=1}^{d} \psi_j(t_j) \int_{\mathbb{R}^d} e^{it\cdot x} d\mu(x) \\ &= \int_{[-T,T]^d} \prod_{j=1}^{d} \psi_j(t_j) \int_{\mathbb{R}^d} e^{it_1\cdot x_1 + \ldots + it_d\cdot X_d} d\mu(x)dt \\ &= \int_{\mathbb{R}^d} \int_{[-T,T]^d} \prod_{j=1}^{d} \psi_j(t_j)e^{it\cdot X} dtd\mu(x) \\ &= \int_{\mathbb{R}^d} \int_{[-T,T]^d} \prod_{j=1}^{d} \psi_j(t_j)e^{it_jX_j} dtd\mu(x) \\ &= \int_{\mathbb{R}^d} \left[\prod_{j=1}^{d} \int_{[-T,T]} \psi_j(t_j)e^{it_jX_j} dt_j\right] d\mu(x) \\ &\to \int_{\mathbb{R}^d} \prod_{j=1}^{d} \left[\pi(1_{(a_j,b_j)}(x_j) + 1_{[a_j,b_j]}(x_j)\right] d\mu(x)\end{aligned}$$

and this proves the result.

Theorem (Continuity Theorem) 8.4. *Let X_n and X be random vectors with characteristic functions φ_n and φ, respectively. Then $X_n \Rightarrow X$ if and only if $\varphi_n(t) \to \varphi(t)$.*

Proof. As before, one direction is trivial. Let $f(x) = e^{itx}$. This is bounded and continuous. $X_n \Rightarrow X$ implies $\varphi_n(x) = E(f(X_n)) \to \varphi(t)$.

For the other direction we need to show tightness. Fix $\theta \in \mathbb{R}^d$. Then for $\forall\ s \in \mathbb{R}$ $\varphi_n(s\theta) \to \varphi(s\theta)$. Let $\tilde{X}_n = \theta \cdot X_n$. Then $\varphi_{\tilde{X}}(s) = \varphi_X(\theta s)$ and

$$\varphi_{\tilde{X}_n}(s) \to \varphi_{\tilde{X}}(s).$$

Therefore the distribution of $\tilde{X}_n$ is tight by what we did earlier. Thus the random variables $e_j \cdot X_n$ are tight. Let $\varepsilon > 0$. There exists a constant positive constant M_i such that

$$\lim_n P(e_j \cdot X_i \in [M_i, M_i]) \geq 1 - \varepsilon.$$

Now take $M = \max_{1 \leq j \leq d} M_j$. Then

$$P(X_n \in [M, M]^d) \geq 1 - \varepsilon$$

and the result follows.

Remark. As before, if $\varphi_n(t) \to \varphi(t)$ and φ is continuous at 0, then $\varphi(t)$ is the characteristic function of a random vector X and $X_n \Rightarrow X$.

Also, it follows from the above argument that If $\theta \cdot X_n \Rightarrow \theta \cdot X$ for all $\theta \in \mathbb{R}^d$ then $X_n \Rightarrow X$. This is often called the **Cramér–Wold devise.**

Next let $X = (X_1, \dots, X_d)$ be independent $X_i \sim N(0,1)$. Then X has density $\frac{1}{(2\pi)^{d/2}} e^{-|x|^2/2}$ where $|x|^2 = \sum_{i=1}^d |x_i|^2$. This is called the standard normal distribution in $\mathbb{R}^d$ and its characteristic function is

$$\varphi_X(t) = E\left(\prod_{j=1}^d e^{it_j X_j}\right) = e^{-|t|^2/2}.$$

Let $A = (a_{ij})$ be a $d \times d$ matrix. and set $Y = AX$ where X is standard normal. The covariance matrix of this new random vector is

$$\begin{aligned}
\Gamma_{ij} &= E(Y_i Y_j) \\
&= E\bigg(\sum_{l=1}^{d} a_{il} X_l \cdot \sum_{m=1}^{d} a_{jm} X_m\bigg) \\
&= \sum_{l=1}^{d} \sum_{m=1}^{d} a_{il} a_{jm} E(X_l X_m) \\
&= \sum_{l=1}^{d} a_{il} a_{jl}.
\end{aligned}$$

Thus $\Gamma = (\Gamma_{ij}) = AA^T$. and the matrix γ is symmetric; $\Gamma^T = \Gamma$. Also the quadratic form of Γ is positive semidefinite. That is,

$$\begin{aligned}
\sum_{ij} \Gamma_{ij} t_i t_j = \langle \Gamma t, t \rangle &= \langle A^T t,\ A^T t \rangle \\
&= |A^T t|^2 \geq 0.
\end{aligned}$$

$$\begin{aligned}
\varphi_Y(t) &= E(e^{it \cdot AX}) \\
&= E(e^{iA^T t \cdot X}) \\
&= e^{-\frac{|A^T t|^2}{2}} \\
&= e^{-\sum\limits_{ij} \Gamma_{ij} t_i t_j}.
\end{aligned}$$

So, the random vector $Y = AX$ has a multivariate normal distribution with covariance matrix Γ.

Conversely, let Γ be a symmetric and nonnegative definite $d \times d$ matrix. Then there exists an orthogonal matrix $\mathcal{O}$ such that

$$O^T \Gamma O = D$$

where D is diagonal. Let $D_0 = \sqrt{D}$ and $A = OD_0$.
Then $AA^T = OD_0(D_0^T O^T) = ODO^T = \Gamma$. So, if we let $Y = AX,\ X$ normal, then Y is multivariate normal with covariance matrix Γ. If Γ is non–singular, so is A and Y has a density.

Theorem 8.5. *Let $X_1, X_2, \ldots$ be i.i.d. random vectors, $EX_n = \mu$ and covariance matrix*

$$\Gamma_{ij} = E(X_{1,j} - \mu_j)(X_{1,i} - \mu_i)).$$

If $S_n = X_1 + \ldots + X_n$ then

$$\frac{S_n - n\mu}{\sqrt{n}} \Rightarrow \chi$$

where χ is a multivariate normal with covariance matrix $\Gamma = (\Gamma_{ij})$.

Proof. By setting $X_n' = X_n - \mu$ we may assume $\mu = 0$. Let $t \in \mathbb{R}^d$. Then $\tilde{X}_n = t \cdot X_n$ are i.i.d. random variables with $E(\tilde{X}_n) = 0$ and

$$E|\tilde{X}_n|^2 = E\left(\sum_{i=1}^{d} t_i (X_n)_i\right)^2 = \sum_{ij} t_i t_j \Gamma_{ij}.$$

So, with $\tilde{S}_n = \sum_{j=1}^{n} (t \cdot X_j)$ we have

$$\varphi_{\tilde{S}_n}(1) = E(e^{i\tilde{S}_n}) \to e^{-\sum\limits_{ij} \Gamma_{ij} t_i t_j / 2}$$

This is equivalent to

$$\varphi_{S_n}(t) = E(e^{it \cdot S_n}) \to e^{-\sum\limits_{ij} \Gamma_{ij} t_i t_j / 2}. \quad \square$$

Theorem. *Let X_i be i.i.d. $E|X_i|^2 = \sigma^2,\ EX_i = 0$ and $E|X_i|^3 = \rho < \infty$. Then if F_n is the distribution of $\dfrac{S_n}{\sigma\sqrt{n}}$ and $\Phi(x)$ is the normal we have*

$$\sup_{x \in \mathbb{R}} |F_n(x) - \Phi(x)| \leq \frac{c\rho}{\sigma^3 \sqrt{n}} \quad (\textit{may take } c = 3).$$

?? is actually true:

$$F_n(x) = \phi(x) + \frac{H_1(x)}{\sqrt{n}} + \frac{H_2(x)}{n} + \ldots + \frac{H_3(x)}{n^{3/2}} + \ldots$$

where $H_i(x)$ are explicit functions involving Hermid <u>*polynomials*</u>*.*

Lemma 1. *Let F be a d.f., G a real–valued function with the following conditions:*

(i) $\lim\limits_{x \to -\infty} G(x) = 0, \ \lim\limits_{x \to +\infty} G(x) = 1,$

(ii) G has bounded derivative with $\sup\limits_{x \in \mathbb{R}} |G'(x)| \leq M$. *Set* $A = \dfrac{1}{2M} \sup\limits_{x \in \mathbb{R}} |F(x) - G(x)|$.

There is a number a s.t. $\forall\ T > 0$

$$\begin{aligned} &2MT\Delta \left\{ 3 \int_0^{T\Delta} \frac{1 - \cos x}{x^2}\ dx - \pi \right\} \\ &\leq \left| \int_{-\infty}^{\infty} \frac{1 - \cos Tx}{x^2} \{F(x+a) - G(x+a)\} dx \right|. \end{aligned}$$

Proof. $\Delta < \infty$ since G is bounded. Assume L.H.S. is > 0 so that $\Delta > 0$. $\dfrac{\frac{b<-a}{a>0}}{\underset{-b<-a}{a < b}}$.

Since $F - G = 0$ at $\pm\infty$, $\exists$ sequence $x_n \to b \in \mathbb{R}$ s.t.

$$F(x_n) - G(x_n) \to \begin{cases} 2M\Delta \\ or \\ -2M\Delta \end{cases}.$$

So, either

$$\begin{cases} F(b) - G(b) = 2M\Delta \\ or \\ F(b-) - G(b) = -2M\Delta. \end{cases}$$

Assume $F(b-) - G(b) = -2M\Delta$.

Put

$$a = b - \Delta < b, \text{ since}$$

$$\Delta = (b - a)$$

if $|x| < \Delta$ we have

$$\begin{aligned} G(b) - G(x+a) &= G'(\xi)(b-a-x) \\ &= G'(\xi)(\Delta - x) \qquad |G'(\xi)| \le M \end{aligned}$$

or

$$\begin{aligned} G(x+a) &= G(b) + (x-\Delta)G'(\xi) \\ &\geq G(b) + (x-\Delta)M. \end{aligned}$$

So that

$$\begin{aligned} F(x+a) - G(x+a) &\leq F(b-) - [G(b) + (x-\Delta)M] \\ &= -2M\Delta - xM + \Delta M \\ &= -M(x+\Delta) \ \forall\ x \in [-\Delta, \Delta] \end{aligned}$$

$\therefore$ T to be chosen: we will consider $\int\limits_{-\infty}^{\infty} = \int\limits_{-\Delta}^{\Delta} + rest$

$$\begin{aligned} &\int_{-\Delta}^{\Delta} \frac{1-\cos Tx}{x^2}\{F(x+a) - G(x+a)\}dx \\ &\leq -M\int_{-\Delta}^{\Delta} \frac{1-\cos Tx}{x^2}(x+\Delta)dx \\ &= -2M\Delta \int_0^{\Delta} \left(\frac{1-\cos Tx}{x^2}\right) dx \end{aligned} \tag{1}$$

$$\begin{aligned} &\left|\left\{\int_{-\infty}^{-\Delta} + \int_{\Delta}^{\infty}\right\}\{F(x+a) - G(x+a)\}dx\right| \\ &\qquad \leq 2M\Delta\left\{\int_{-\infty}^{-\Delta} + \int_{\Delta}^{\infty}\right\}\frac{1+\cos Tx}{x^2}dx = 4M\Delta\int_{\Delta}^{\infty} \frac{1-\cos Tx}{x^2}\ dx. \end{aligned} \tag{2}$$

Add:

$$
\begin{aligned}
&\int_{-\infty}^{\infty}\left(\frac{1-\cos Tx}{x^2}\right)\{F(x+a)-G(x+a)\}dx \\
&\leq 2M\Delta\left\{-\int_0^{\Delta}+2\int_{\Delta}^{\infty}\right\}\left\{\frac{1-\cos Tx}{x^2}\right\}dx \\
&= 2M\Delta\left\{-3\int_0^{\Delta}+2\int_0^{\infty}\right\}\left\{\frac{1-\cos Tx}{x^2}\right\}dx \\
&= 2M\Delta\left\{-3\int_0^{\Delta}\frac{1-\cos Tx}{x^2}dx+2\int_0^{\infty}\frac{1-\cos Tx}{x^2}\ dx\right\} \\
&= 2M\Delta\left\{-3\int_0^{\Delta}\frac{1-\cos Tx}{x^2}dx+2\left(\frac{\pi T}{2}\right)\right\} \\
&= 2MT\Delta\left\{-3\int_0^{T\Delta}\frac{1-\cos x}{x^2}dx+\pi\right\}<0.
\end{aligned}
$$

Lemma 2. *Suppose in addition that*

(iii) G is of bounded variation in $(-\infty,\infty)$. (Assume G has a density).

(iv) $\displaystyle\int\limits_{-\infty}^{\infty}|F(x)-G(x)|dx<\infty.$

Let

$$f(t)=\int_{-\infty}^{\infty}e^{itx}dF(x),\ \ g(t)=\int_{-\infty}^{\infty}e^{itx}dG(x).$$

Then

$$\Delta\leq\frac{1}{2\pi M}\int_{-T}^{T}\frac{|f(t)-g(t)|}{t}dt+\frac{12}{\pi T}.$$

Proof.

$$f(t)-g(t)=-it\int_{-\infty}^{\infty}\{F(x)-G(x)\}e^{itx}dx$$

and $\therefore$

$$
\begin{aligned}
\frac{f(t)-g(t)}{-it}\ e^{-ita} &= \int_{-\infty}^{\infty}(F(x)-G(x))e^{-ita+itx}dx \\
&= \int_{-\infty}^{\infty}F(x+a)-G(x+a)e^{+itx}dx.
\end{aligned}
$$

$\therefore$ By (iv), R.H.S. is bounded and L.H.S. is also. Multiply left hand side by $(T-|t|)$ and integrade

$$\begin{aligned}
&\int_{-T}^{T}\left\{\frac{f(t)-g(t)}{-it}\right\}e^{ita}(T-|t|)dt\\
&\quad=\int_{-T}^{T}\int_{-\infty}^{\infty}\{F(x+a)-G(x+a)\}e^{itx}(T-|t|)dxdt\\
&\quad=\int_{-\infty}^{\infty}\{F(x+a)-G(x+a)\}\int_{-T}^{T}e^{itx}(T-|t|)dtdx\\
&\quad=\int_{-\infty}^{\infty}(F(x+a)-G(x+a))\int_{-T}^{T}e^{itx}(T-|t|)dtdx
\end{aligned}$$

Now,

$$\begin{aligned}
&\frac{1-\cos Tx}{x^2}=\frac{1}{2}\int_{-T}^{T}(T-|t|)e^{itx}dt\\
&\overset{\text{above}}{=}2\int_{-\infty}^{\infty}(F(x+a)-G(x+a))\left\{\frac{1-\cos Tx}{x^2}\right\}dx\\
&\text{or }\left|\int_{-\infty}^{\infty}\Big\{F(x+a)-G(x+a)\Big\}\left\{\frac{1-\cos Tx}{x^2}\right\}dx\right|\\
&\quad\leq\frac{1}{2}\left|\int_{-T}^{T}\frac{f(t)-g(t)}{-it}\,e^{-ita}(T-|t|)dt\right|\\
&\quad\leq T/2\int_{-T}^{T}\left|\frac{f(t)-g(t)}{t}\right|dt
\end{aligned}$$

$\therefore$ Lemma 1 $\Rightarrow$

$$2M\Delta\left\{3\int_0^{T\Delta}\frac{1-\cos x}{x^2}\,dx-\pi\right\}\leq\int_{-T}^{T}\frac{|f(t)-g(t)|}{t}\,dt$$

or now,

$$\begin{aligned}
3\int_0^{T\Delta'}\frac{1-\cos x}{x^2}\,dx-\pi&=3\int_0^{\infty}\frac{1-\cos x}{x^2}\,dx-3\int_{T\Delta}^{\infty}\frac{1-\cos x}{x^2}\,dx-\pi\\
&=3\left(\frac{\pi}{2}\right)-3\int_{T\Delta}^{\infty}\frac{1-\cos x}{x^2}\,dx-\pi\\
&\geq\frac{3\pi}{2}-6\int_{T\Delta}^{\infty}\frac{dx}{x^2}-\pi=\frac{\pi}{2}-\frac{6}{T\Delta}
\end{aligned}$$

or

$$\int_{-T}^{T} \frac{|f(t)-g(t)|}{t}\,dt \geq 2\left(2M\Delta\left(\frac{\pi}{2}-\frac{6}{T\Delta}\right)\right)$$
$$= 2M\pi\Delta - \frac{24M}{T}$$

or

$$\Delta \leq \frac{1}{2MT}\int_{-T}^{T} \frac{|f(t)-g(t)|}{|t|}\,dt + \frac{12}{\pi}$$

We have now bound the difference between the d.f. satisfying certain conditions by the average difference of.

Now we apply this to our functions: Assume $w \log \sigma^2 = 1,\ \rho \geq 1.$

$$F(x) = F_n(x) = P\left(\frac{S_n}{\sqrt{n}} > x\right),\ X_i i.i.d.$$
$$G(x) = \phi(x) = P(N > x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-y^2/2}dy.$$

Clearly F_n and Φ satisfy (i).

$$\sup_{x\in} |\Phi'(x)| = \frac{1}{\sqrt{2\pi}}e^{-x^2/2} = \frac{1}{\sqrt{2\pi}} = .39894 < 2/5 = M.$$

(iii) Satisfy:

(iv) $\int\limits_{\mathbb{R}} |F_n(x) - \Phi(x)|dx < \infty.$

Clearly, $\int_{-1}^{1} |F_n(x) - \phi(x)|dx < \infty.$

Need:

$$\int_{-\infty}^{-1} |F_n(x) - \Phi(x)|dx + \int_{1}^{\infty} |F(x) - \Phi(x)|dx < \infty.$$

Assume $w \log \sigma^2 = 1$. For $x > 0$. $P(|X| > x) = \frac{1}{\lambda^2}F|x|^2$.

$$1 - F_n(x) = P\left(\frac{S_n}{\sqrt{n}} > x\right) \leq \frac{1}{|X|^2}E\left|\frac{S_n}{\sqrt{n}}\right|^2 < \frac{1}{|X|^2}$$
$$1 - \Phi(x) = P(N > x) \leq \frac{1}{x^2}E|N|^2 = \frac{1}{|x|^2}.$$

In particular: for $x > 0$. $\max((1 - F_n(x)),\ \max 1 - \Phi(x)) \leq \frac{1}{|x|^2}$.

If $x < 0$. Then

$$F_n(x) = P\left(\frac{S_n}{\sqrt{n}} < x\right) = P\left(-\frac{S_n}{\sqrt{n}} > -x\right) \leq \frac{1}{x^2} E\left|\frac{S_n}{\sqrt{n}}\right|^2 = \frac{1}{x^2}$$

$$\Phi(x) = P(N < x) \leq \frac{1}{x^2}$$

$$\therefore \overset{x<0}{\max}(F_n(x),\ \Phi(x)) \leq \frac{1}{x^2}$$

$$|F(x) - \Phi(x)| \leq \frac{1}{x^2} \text{ if } x < 0$$

$$\{|F(x) - \phi(x)| =$$

$$= |1 - \phi(x) - (1 - F(x))| \leq \frac{1}{x^2}. \qquad x > 0$$

$\therefore$ (iv) hold.

$$\begin{aligned} |F_n(x) - \Phi(x)| &\leq \frac{1}{\pi} \int_{-T}^{T} \frac{|\varphi^n(t/\sqrt{n}) - e^{-t^2/2}|}{|t|}\ dt + \frac{24M}{\pi T} \\ &\leq \frac{1}{\pi} \int_{-T}^{T} \frac{|\varphi^n(t/\sqrt{n}) - e^{-t^2/2}}{|t|}\ dt + \frac{48}{5\pi T} \end{aligned}$$

tells us what we must do. Take $T =$ multiple of $\sqrt{n}$.

Assume $n \geq 10$.

Take $T = \frac{4\sqrt{n}}{3\rho}$: Then

$$\frac{48}{5\pi T} = \frac{48 \cdot 3}{5\pi 4\sqrt{n}}\rho = \frac{12 \cdot 3}{5\pi\sqrt{n}}\ \rho = \boxed{\frac{36\rho}{5\pi\sqrt{n}}}.$$

For second. Claim:

$$\frac{1}{|t|}|\varphi^n(t/\sqrt{n}) - e^{-t^2/2}| \leq$$

$$\leq \frac{1}{T} e^{-t^2/4} \left\{\frac{2t^2}{9} + \frac{|t|^3}{18}\right\},\ \begin{matrix} -T \leq t \leq T \\ T = 4\sqrt{n}/3\rho \end{matrix}, \quad n \geq 10$$

$$\therefore\ \pi T|F_n(x)-\Phi(x)| \le \int_{-T}^{T} e^{-t^2/4}\left\{\frac{2t^2}{9}+\frac{|t|^3}{18}\right\}dt$$

$$+\frac{48}{5}-$$

$$= \int_{-T}^{T} e^{-t^2/4}\left\{\frac{2t^2}{9}+\frac{|t|^3}{18}\right\}dt + \overset{9.6}{\frac{48}{5}}$$

$$\le \frac{2}{9}\int_{-\infty}^{\infty} e^{-t^2/4}t^2dt + \frac{1}{18}\int_{-\infty}^{\infty} e^{-t^2/4}|t|^3dt + 9.6$$

$$= I + II + 9.6.$$

Recall: $\displaystyle\frac{1}{\sqrt{2\pi\sigma^2}}\int e^{-t^2/2\sigma^2}t^2dt = \sigma^2.$

Take $\sigma^2 = 2$

$$\frac{2}{9}\int_{-\infty}^{\infty} e^{-t^2/2\sigma^2}t^2dt = \frac{2}{9}\sqrt{2\pi\cdot 2}\cdot 2 = \frac{2\cdot 2\cdot 2}{9}\sqrt{\pi}$$

$$= \frac{8}{9}\sqrt{\pi}$$

$$\frac{1}{18}\left\{\int_{-\infty}^{\infty}|t|^3e^{-t^2/2\sigma^2}dt\right\}$$

$$= \frac{1}{18}\left\{2\int_0^{\infty}t^3e^{-t^2/4}dt = \frac{1}{18}\left\{2\int_0^{\infty}t^2\cdot te^{-t^2/4}dt\right\}\right.$$

$$= \frac{1}{18}+\left\{2\int_0^{\infty}2t\cdot 2e^{-t^2/4}dt\right\}$$

$$= \frac{1}{18}\left\{8\int_0^{\infty}te^{-t^2/4}dt = \left\{-16e^{-t^2/4}\Big|_0^{\infty}\right\}\frac{1}{18}\right.$$

$$= \frac{16}{18} = \frac{8}{9}$$

$$\pi T|F_n(x)-\Phi(x)| \le \left(\frac{8}{9}\sqrt{\pi}+\frac{8}{9}+9.6\right).$$

or

$$|F_n(x)-\Phi(x)| \le \frac{1}{\pi T}\left\{\frac{8}{9}(1+\sqrt{\pi}\Big)+9.6\right\}$$

$$\le \frac{3\rho}{4\sqrt{n}\ \pi}\left\{\frac{8}{9}(1+\sqrt{\pi})+9.6\right\}$$

$$= \frac{\rho}{\sqrt{n}}\left\{\underbrace{\qquad}\right\} < \frac{3\rho}{\sqrt{n}}.$$

For $n \le 9$, the result follows. $\boxed{1 < \rho}$ since $\sigma^2 = 1$.

Proof of claim. Recall: $\sigma^2 = 1$, $\left|\varphi(t) - \sum_{m=0}^{n} \frac{E(itX)^m}{m!}\right| \le E(\min \frac{|t|X|^{nt}}{(n+1)!}\ \frac{2|tX|^n}{n!}\Big)$.

(1) $\left|\varphi(t) - 1 + \frac{t^2}{2}\right| \le \frac{\rho|t|^3}{6}$ and

(2) $|\varphi(t)| \le 1 - t^2/2 + \frac{\rho|t|^3}{6}$, for $t^2 \le 2$.

So, if $T = \frac{4\sqrt{n}}{3\rho}$ if $|t| \le L \Rightarrow \left(\frac{\rho|t|}{\sqrt{n}}\right) \le (4/3) = \frac{16}{9} < 2$.

$\Rightarrow$ Also, $t/\sqrt{n} = \frac{4}{3\rho} < 2$. So,

$$\begin{aligned}
\left|\varphi\left(\frac{t}{\sqrt{n}}\right)\right| &\le 1 - \frac{t^2}{2n} + \frac{\rho|t|^3}{6n^{3/2}} = 1 - \frac{t^2}{2n} + \frac{\rho|t|}{\sqrt{n}}\frac{|t|^2}{n} \\
&\le 1 - \frac{t^2}{2n} + \frac{4}{3}\frac{t^2}{n} = 1 - \frac{5t^2}{18n} \\
&\le e^{\frac{-5t^2}{18n}},\ \text{gives that } 1 - x \le e^{-x}.
\end{aligned}$$

Now, let $\alpha = \varphi(t/\sqrt{n})$, $\beta = e^{-t^2/2n}$ and $\gamma = e^{\frac{-5t^2}{18n}}$. Then $n \ge 10 \Rightarrow \gamma^{n-1} \le e^{-t^2/4}$.

$$\begin{aligned}
&|\alpha^n - \beta^n| \le n\gamma^{n-1}|\alpha - \beta| \Leftrightarrow \\
&|\varphi(t/\sqrt{n}) - e^{-t^2/2}| \le n e^{\frac{-5t^2}{18n}(n-1)}|\varphi(t/\sqrt{n}) - e^{-t^2/2n}| \\
&\le n e^{-t^2/4}|\varphi(t/\sqrt{n}) - 1 + \frac{t^2}{2n} - e^{-t^2/2n - 1 - t^2/2n}| \\
&\le n e^{-t^2/4}|\varphi(t/\sqrt{n}) - 1 + \frac{t^2}{2n}| + n e^{-t^2/4}|1 - \frac{t^2}{2n} - e^{-t^2/2n}| \\
&\le n e^{-t^2/4}\frac{\rho|t|^3}{6n^{3/2}} + n e^{-t^2/4}\frac{t^4}{2 \cdot 4n^2}
\end{aligned}$$

using $|e^{-x} - (1-x)| \le \frac{x^2}{2}$ for $0 < x < 1$

or

$$\begin{aligned}
\frac{1}{|t|}|\varphi(t/\sqrt{n}) - e^{-t^2/2}| &\le \frac{\rho t^2 e^{-t^2/4}}{6\sqrt{n}} + \frac{e^{-t^2/4}|t|^3}{8n} \\
&= e^{-t^2/4}\left\{\frac{\rho t^2}{6\sqrt{n}} + \frac{|t|^3}{8n}\right\} \\
&\le \frac{1}{T}e^{-t^2/4}\left\{\frac{2t^2}{9} + \frac{|t|^3}{18}\right\}
\end{aligned}$$

using $\rho/\sqrt{n} = \frac{4}{3}T$ and $\dfrac{1}{n} = \dfrac{1}{\sqrt{n}}\dfrac{1}{\sqrt{n}} \le \dfrac{4}{3}T \cdot \dfrac{1}{3},\ \rho > 1$ and $n \ge 10$. Q.E.D.

Question: Suppose F has density f. Is it true that the density of $\dfrac{S_n}{\sqrt{n}}$ tends to the density of the normal? This is not always true. (Feller v. 2, p. 489). However, it is true if more conditions.

Let X_i be i.i.d. $EX_i = 0,\ FX_i^2 = 1$.

Theorem. *If $\varphi \in L^1$, then $\dfrac{S_n}{\sqrt{n}}$ has a density f_n which converges uniformly to $\dfrac{1}{\sqrt{2\pi}}\, e^{-x^2/2} = \eta(x)$.*

Proof.

$$
\begin{aligned}
f_n(x) &= \frac{1}{2\pi}\int_{\mathbb{R}} e^{itx}\varphi_n(t)dt \\
n(x) &= \frac{1}{2\pi}\int_{\mathbb{R}} e^{ity}e^{-\frac{1}{2}t^2}\,dt \\
\therefore\ |f_n(x) - \eta(x)| &\le \frac{1}{2\pi}\int_{-\infty}^{\infty} |\varphi(t/\sqrt{n})^n - e^{-1/2t^2}|dt
\end{aligned}
$$

under the assumption

$$|\varphi(t)| \le e^{-\frac{1}{4}t^2} \text{ for } |t| < \delta.$$

At 0, both sides are 0.

Both have ?? derivatives of social derivative of $\varphi(t)$ at 0 is -1 smaller than the second derivative of r.h.s.

Limit Theorems in $\mathbb{R}^d$

Recall: $\mathbb{R}^d = \{(x_1, \dots, x_d) : x_i \in \mathbb{R}\}$.
If $X = (x_1, \dots, X_d)$ is a random vector, i.e. a r.v. $X : \omega \to \mathbb{R}^d$. We defined its distribution function by $F(x) = P(X \le x)$, where $X \le x \Leftrightarrow X_i \le x_i,\ i = 1, \dots, d$. F has the following properties:

(i) $x \le y \Rightarrow F(x) \le F(y)$.

(ii) $\lim\limits_{x\to\infty} F(x) = 1, \ \lim\limits_{x_i\to-\infty} F(x) = 0$.

(iii) F is right cont. i.e. $\lim\limits_{y\downarrow x} F(x) = F(x)$.

$X_p \to \infty$ we mean each coordinate goes to zero. You know what $X_i \to -\infty$.

There is also the distribution measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$: $\mu(A) = P(X \in A)$.

If you have a function satisfying (i) $\leftrightarrow$ (ii), this may not induce a measure. Example: we must have:

$$
\begin{aligned}
&P(X \in (a_1, b_1] \times (a_2, b_2]) = F(b_1, b_2) - F(a_1, b_2) \\
&P(a < X_1 \le b_1, a_2 \le X_2 \le b_2) - F(b_1, a_2) + F(a_1, a_2).
\end{aligned}
$$

Need: measure of each vect. ≥ 0,

Example. $f(x_1, x_2) = \begin{cases} 1 & x_1, x_1 \ge 1 \\ 2/3 & \\ 2/3 & x_1 \ge 1, \ 0 \le x_2 \le 1 \\ 0 & x_2 \ge 1, \ 0 \le x_1 < 1 \\ & \text{else} \end{cases}$.

If $0 < a_1, \ a_2 < 1 \le b_1, \ b_2 < \infty \Rightarrow$

$$
\begin{aligned}
F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) &= 1 - 2/3 - 2/3 + 0 \\
&= -1/3.
\end{aligned}
$$

The measure has:

$$\boxed{\mu(0,1) = \mu(1,0) = 2/3,\ \mu(1,1) = -1/3}$$

for each, need measure of ??

Other simple ??

Recall: If F is the dist. of $(x_1, \dots, X_n)$, then $F_i(x) = P(X_i \leq x)$, x real in the marginal distributions of F

$$F_i(x) = \lim_{m \to \infty} F(m, \dots, m, x_{i+1}, \dots m)$$

F has a density if $\exists\ f \geq 0$ with $\displaystyle\int_{\mathbb{R}^d} 0 = f$ and

$$F(x_1, x_2, x_d) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f(y) dy_1 \dots dy_2.$$

<u>Def</u>: If $F\ F_n$ is a distribution function in $\mathbb{R}^d$, we say F_n converges weakly to F, if $F_n \Rightarrow F$, if $\lim_{n \to \infty} F_n(x) = F(x)$ for all pts of continuity of F.
$X_n \Rightarrow X,\ \mu_n \Rightarrow \mu.$

<u>Recall</u>: $\overline{A}$ = set of limits of sequences in A, closure of A. $A^o = \mathbb{R}^d \backslash \overline{(\mathbb{R}^d | A)}$ <u>interior</u>. $\partial A = \overline{A} - A^o$. A Borel set A is a μ–continuity set if $\mu(\partial A) = 0$.

Theorem 1 Skorho**.** *$X_n \Rightarrow X_\infty \Rightarrow \exists$ r.v. $X_n \sim X_n,\ Y \sim X_\infty$ s.t. $Y_n \to Y$ a.e.*

Theorem 2. *The following statements are equivalent to $X_n \Rightarrow X_\infty$.*

(i) $Ef(X_n) \to E(f(X_\infty)) \forall$ *bounded cont.* f.

(iii) $\forall$ *closed sets* k, $\overline{\lim} P(X_n \in k) \leq P(X_\infty \in k)$.

(iv) $\forall$ *open sets* G, $\underline{\lim} P(X_n \in G) \geq P(X_\infty \in G)$.

(v) $\forall$ *continuity* A, $(P(X \in \partial A)$.

$$\lim_{n \to \infty} P(X_n \in A) = P(X_\infty \in A).$$

(vi) *Let D_p = discontinuity sets of f. If $P(X_\infty \in D_p) = 0 \Rightarrow E(f(X_n)) \to E(f(X_\infty))$, f bounded.*

Proof. $X_n \Rightarrow X_\infty \Rightarrow$ (i) trivial. (i) $\Rightarrow$ (ii) trivial. (i) $\Rightarrow$ (ii). Let $d(x,k) = \inf\{d(x-y) : y \in k\}$.

$$\varphi_j(t) = \begin{cases} 1 & t \le 0 \\ 1 - jt & 0 \le t \le j^{-1} \\ 0 & {}^{-1} \le t \end{cases}$$

Let $f_j(x) = \varphi_j(\text{dist }(x,k))$. f_j is cont. and bounded by 1 and $f_j(x) \downarrow I_k(x)$ since k is closed.

$$\begin{aligned} \therefore \limsup_{n\to\infty} \mu_n(k) &= \lim_{n\to\infty} E(f_j(X_n)) \\ &= E(f_j(X_\infty)) \downarrow P(X_\infty \in k). \qquad \text{Q.E.D.} \end{aligned}$$

(iii) $\Rightarrow$ (iv): A open iff A^c closed and $P(X \in A) + P(X \in A^c) = 1$.

(v) $\Rightarrow$ implies conv. in dis. If F is cont. at X, then with $A = (-\infty, x_1] \times \dots (-\infty, x_d]$, $x = (x_1, \dots, x_n)$ we have $\mu(\partial A) = 0$. So, $F_n(x) = F(x_n \in A) \to P(X_\infty \in A) = F(x)$. Q.E.D.

As in 1–dim. μ_n is tight if $\exists$ given $\varepsilon \ge 0$ s.t.

$$\inf_n \mu_n([-M,M]^d) \ge 1 - \varepsilon.$$

Theorem. *If μ_n is tight $\Rightarrow \exists \mu_{n_j}$ s.t. $\mu_{n_j} \xrightarrow{\text{weak}} \mu$.*

Ch. f. Let $X = (X_1, \dots, X_d)$ be a r. vector. $\varphi(t) = E(e^{it\cdot X})$ is the Ch.f. $t \cdot X = t_1 X_1 + \ldots + t_d X_d$.

Inversion formula: Let $A = [a_1, b_1] \times \ldots \times [a_d, b_d]$ with $\mu(\partial A) = 0$. Then

$$\mu(A) = \lim_{T\to\infty} \frac{1}{(2\pi)^d} \int_{-T}^{T} \cdots \int_{-T}^{T} \psi_1(t_1)\varphi(t) \ldots \psi_d(t_d)\varphi(t) dt_1, \dots, dt_2$$

where

$$\psi_j(s) = \left(\frac{e^{isa_j} - e^{-sb_j}}{is} \right)$$
$$= \lim_{T\to\infty} \frac{1}{(2\pi)^d} \int_{[-T,T]^d} \prod_{j=1}^{d} \psi_j(t_j)\varphi(t)dt$$

Proof. Apply Fubini's Theorem:

$A = [a_1, b_1] \times \ldots \times [a_d, b_d]$ with $\mu(\partial A) = 0$. Then

$$\int_{[-T,T]^d} \prod_{j=1}^{d} \psi_j(t_j) \int_{\mathbb{R}^d} e^{it\cdot x} d\mu(x)$$
$$= \int_{[-T,T]^d} \prod_{j=1}^{d} \psi_j(t_j) \int_{\mathbb{R}^d} e^{it_1\cdot x_1 + \ldots + it_d \cdot X_d} d\mu(x)dt$$
$$= \int_{\mathbb{R}^d} \int [-T,T]^d \prod_{j=1}^{d} \psi_j(t_j) e^{it\cdot X} dt d\mu(x)$$
$$= \int_{\mathbb{R}^d} \int_{[-T,T]^d} \prod_{j=1}^{d} \psi_j(t_j) e^{itjX_j} dt d\mu(x)$$
$$= \int_{\mathbb{R}^d} \left[\prod_{j=1}^{d} \int_{[-T,T]} \psi_j(t_j) e^{itjX_j} dt_j \right] d\mu(x)$$
$$\to \int_{\mathbb{R}^d} \prod_{j=1}^{d} \left[\pi(1_{(a_j,b_j)}(x_j) + 1_{[a_j,b_j]}(x_j) \right] d\mu(x)$$

results $= \mu(A)$.

Continuity Theorem. *Let $X_n,\ 1 \le n \le \infty$ be random vectors with Ch. f.'s φ_n. Then $X_n \Rightarrow X_\infty \Leftrightarrow \varphi_n(t) \to \varphi_\infty(t)$.*

Proof. $f(x) = e^{itx}$ is bounded and cont. to, $X_n \Rightarrow x$ implies $\varphi_n(x) = E(f(X_n)) \to \varphi_\infty(t)$.

Next, we show as earlier, that sequence is tight.

Fix $\mathcal{O} \in \mathbb{R}^d$. Then for $\forall\ s \in \mathbb{R}\ \varphi_n(s\mathcal{O}) \to \varphi_\infty(s\mathcal{O})$. Let $\tilde{X}_n = \mathcal{O} \cdot X_n$: Then $\varphi_{\tilde{X}}(s) = \varphi_X(\mathcal{O}s)$. Then

$$\varphi_{\tilde{X}_n}(s) \to \varphi_{\tilde{X}}(s).$$

$\therefore$ The dist. of $\tilde{X}_n$ is tight.

Thus, for each vector, $e_j \cdot X_n$ is tight.

So, given $\varepsilon > 0,\ \exists M_i$ s.t.

$$\lim_n P(X_i^n \in [M_i, M_i]) > 1 - \varepsilon.$$

Now, $\pi[M_i, M_i]$ take $M =$ largest of all.

$$P(X_n \in [M, M]) \geq 1 - \varepsilon. \qquad \text{Q.E.D.}$$

Remark. As before, if $\varphi_n(t) \to \varphi_a(t)$ and φ_∞ is cont. at 0, then $\varphi_\infty(t)$ is the Ch.f. of a r.vec. X_∞ and $X_n \Rightarrow X_\infty$.

<u>We also showed: Cramér–Wold device</u>

If $\mathcal{O} \cdot X_n \Rightarrow \mathcal{O} \cdot X_\infty \forall\ \mathcal{O} \in \mathbb{R}^d \Rightarrow X_n \Rightarrow X$.

Proof the condition implies $E(e^{i\mathcal{O}\cdot X_n}) \to E(e^{i\mathcal{O}\cdot X_n}) \forall\ \mathcal{O} \in \mathbb{R}^d \cdot \dfrac{\varphi_n(\mathcal{O}) \to \varphi(\mathcal{O})}{\text{Q.E.D.}}$.

Last time: (Continuity Theorem): $X_n \Rightarrow X_\infty$ iff $\varphi_n(t) \to \varphi_\infty(t)$.

We showed: $\mathcal{O} \cdot X_n \Rightarrow \mathcal{O} \cdot X_\infty \forall\ \mathcal{O} \in \mathbb{R}^d$.

Implies: $X_n \Rightarrow X$

This is called Cramér–Wold device.

Next let $X = (X_1, \ldots, X_d)$ be independent $X_i \sim N(0, 1)$. Then X_i has density $\dfrac{1}{\sqrt{2\pi}}\ e^{-\frac{|x|^2}{2\pi}}$.

$\therefore$ X has density

$$\frac{1}{(2\pi)^{d/2}} e^{-|x|^2/2},$$

$$x = (x_1, \dots, x_d), \ |x|^2 = \sum_{i=1}^{d} |x_i|^2.$$

This is called the standard normal.

The Ch.f.

$$\varphi(t) = E\left(\prod_{j=1}^{d} e^{it_j X_j}\right) = e^{-|t|^2/2}.$$

Let $A = (a_{ij})$ be a $d \times d$ matrix. Let

$$\begin{aligned} Y &= AX, \ X \text{ normal} \\ Y_j &= \sum_{l=1}^{d} a_{jl} X_l \end{aligned}$$

Let

$$\begin{aligned} \Gamma_{ij} &= E(Y_i Y_j) \\ &= E\left(\sum_{l=1}^{d} a_{il} X_l \cdot \sum_{m=1}^{d} a_{jm} X_m\right) \\ &= \sum_{l=1}^{d} \sum_{m=1}^{d} a_{il} a_{jm} E(X_l X_m) \\ &= \sum_{l=1}^{d} a_{il} a_{jl} \\ \Gamma = (\Gamma_{ij}) &= AA^T. \end{aligned}$$

Recall: For any matrix, $\langle Bx, x\rangle = \langle x, B^+ x\rangle$. So, Γ is symmetric. $\Gamma^T = \Gamma$.

Also,

$$\begin{aligned} \sum_{ij} \Gamma_{ij} t_i t_j = \langle \Gamma t, t\rangle &= \langle A^t t, \ A^t t\rangle \\ &= |A^t t| \geq 0. \end{aligned}$$

So, Γ is nonnegative definite.

$$\begin{aligned} E(e^{it\cdot AX}) &= E(e^{itA^{\ddagger}t\cdot X}) \\ &= e^{-\frac{|A^T t|^2}{2}} \\ &= e^{-\sum\limits_{ij}\Gamma_{ij}t_i t_j}. \end{aligned}$$

So, the random vector $Y = AX$ has a multivariate normal distribution with covariance matrix Γ.

Conversely: Let Γ be a symmetric and nonnegative definite $d \times d$ matrix. Then $\exists$ $\mathcal{O}$ orthogonal s.t.

$$O^T \Gamma O = D - D \text{ diagonal.}$$

Let $D_0 = \sqrt{D}$ and $A = OD_0$.
Then $AA^T = OD_0(D_0^T O^T) = ODO^T = \Gamma$. So, if we let $Y = AX,\ X$ normal, then Y is multivariate normal with covariance matrix Γ.
If Γ is non–singular, so is A and Y has a density.

Theorem. *Let $X_1, X_2, \ldots$ be i.i.d. random vectors, $EX_n = \mu$ and covariance matrix*

$$\Gamma_{ij} = E(X_{1,j} - \mu_j)(X_{1,i} - \mu_i)).$$

If $S_n = X_1 + \ldots X_n$ then

$$\frac{S_n - n\mu}{\sqrt{n}} \Rightarrow \chi \quad \textit{where}$$

χ is a multivariate normal with covariance matrix $\Gamma = (\Gamma_{ij})$.

Proof. Letting $X_n' = X_n - \mu$ we may assume $\mu = 0$. Let $t \in \mathbb{R}^d$. Then $\tilde{X}_n = t \cdot X_n$ are i.i.d. random variables with $E(\tilde{X}_n) = 0$ and

$$E|\tilde{X}_n|^2 = E\left(\sum_{i=1}^{d} t_i X_{ni}\right)^2 = \sum_{ij} t_i t_j \Gamma_{ij}.$$

So, with $\tilde{S}_n = \sum_{j=1}^{n} (t \cdot X_j)$ we have

$$\varphi_{\tilde{S}_n}(1) = E(e^{i\tilde{S}_n}) \to e^{-\sum\limits_{ij} \Gamma_{ij} t_i t_j / 2}$$

or

$$\varphi_{S_n}(t) = E(e^{it \cdot S_n}) \to \qquad \text{Q.E.D.}$$

Math/Stat 539. Ideas for some of the problems in final homework assignment. Fall 1996.

#1b) Approximate the integral by

$$\frac{1}{n}\sum_{k=1}^{n}\left(\frac{B_k}{n}-\frac{B_{k-1}}{n}\right) \quad \text{and}\ldots$$

$$\begin{aligned} E\left(\int_0^1 B_t dt\right)^2 &= E\left(\int_0^1\int_0^1 B_s B_t ds dt\right) \\ &= 2\int_0^1\int_s^1 E(B_s B_t) dt ds \\ &= 2\int_0^1\int_s^1 s dt ds = \frac{1}{3}. \end{aligned}$$

#2a) Use the estimate $C_p = \dfrac{2^p C_1 e^{c/\varepsilon^2}}{\varepsilon^p(e^{c/\varepsilon^2}-2^p)}$ from class and choose $\varepsilon \sim c/\sqrt{p}$.

#2b) Use (a) and sum the series for the exponential.

#2c) Show $\phi(2\lambda) \le c\phi(\lambda)$ for some constant c. Use formula $E(\phi(X)) = \int\limits_0^\infty \phi'(\lambda) P\{X \ge \lambda\} d\lambda$ and apply good–λ inequalities.

#3a) Use the "exponential" martingale and ...

#3b) Take $b = 0$ in #3a).

#4)(i) As in the proof of the reflection property. Let

$$Y_s^1(\omega) = \begin{cases} 1, & s < t \text{ and } u < \omega(t-s) < v \\ 0 & \text{else} \end{cases}$$

and

$$Y_s^2(\omega) = \begin{cases} 1, & s < t,\ 2a - v < \omega(t-s) < 2a - u \\ 0 & \text{else.} \end{cases}$$

Then $E_x(Y_s^1) = E_x(Y_s^2)$ (why?) and with $\tau = (\inf\{s\colon B_s = a\}) \wedge t$, we apply the strong Markov property to get

$$E_x(Y_\tau^1 \circ \theta_\tau | \mathcal{F}_\tau) \overset{\text{why}}{=} E_x(Y_\tau^2 \circ \theta_\tau | \mathcal{F}_\tau)$$

and ... gives the result.

#4)(i) Let the interval $(u, v) \downarrow x$ to get

$$P_0\{M_t > a,\ B_t = x\} = P_0\{B_t = 2a - x\} = \frac{1}{\sqrt{2\pi t}}\ e^{-\frac{(2a-x)^2}{2t}}$$

and differentiate with respect to a.

#5a) Follow Durrett, page 402, and apply the Markov property at the end.

#7)(i)

$$\begin{aligned} E(X_{n+1}|\mathcal{F}_n) &= e^{\theta S_n - (n+1)\psi(\theta)} E(e^{\theta \xi_{n+1}}|\mathcal{F}_n) \\ &= e^{\theta S_n - n\psi(\theta)} (\xi_{n+1} \text{ independent of } \mathcal{F}_n). \end{aligned}$$

(ii) Show $\psi'(\theta) = \varphi'(\theta)/\varphi(\theta)$ and

$$\left(\frac{\varphi'(\theta)}{\varphi(\theta)}\right) = \frac{\varphi''(\theta)}{\varphi(\theta)} - \left(\frac{\varphi'(\theta)}{\varphi(\theta)}\right)^2 = E(Y_\theta^2) - (E(Y_\theta))^2 > 0$$

where Y_θ has distribution $\dfrac{e^{\theta x}}{\varphi(\theta)}$ (distribution of ξ_1). (Why is true?)

(iii)

$$\begin{aligned} \sqrt{X_n^\theta} &= e^{\frac{\theta}{2}\ S_n - \frac{n}{2}\ \psi(\theta)} \\ &= X_n^{\theta/2}\ e^{n\{\psi(\frac{\theta}{2}) - \frac{1}{2}\psi(\theta)\}} \end{aligned}$$

Strict convexity, $\psi(0) = 0$, and ... imply that

$$E\sqrt{X_n^\theta} = e^{n\{\psi(\frac{\theta}{2}) - \frac{1}{2}\psi(\theta)\}} \to 0$$

as $n \to \infty$. This implies $X_n^\theta \to 0$ in probability.

Chapter 7

1) Conditional Expectation.

   (a) The Radon–Nikodym Theorem.

Durrett p. 476

Signed measures: If μ_1 and μ_2 are two measures, particularly prob. measures, we could add them. i.e. $\mu = \mu_1 + \mu_2$ is a measure.

But what about $\mu_1 - \mu_2$?

Definition 1.1. By a signed measure on a measurable space $(\Omega, \mathcal{F})$ we mean an extended real valued function ν defined on $\mathcal{F}$ such that

(i) ν assumes at most one of the values $+\infty$ or $-\infty$.

(ii) $\nu\left(\bigcup_{j=1}^{\infty} E_j\right) = \sum_{j=1}^{\infty} \nu(E_j)$, E_j's are disjoint in $\mathcal{F}$. By (iii) we mean that the series is absolutely convergent if $\nu\left(\bigcup_{j=1}^{\infty} E_j\right)$ is finite and properly divergent if $\nu\left(\bigcup_{j=1}^{\infty} E_j\right)$ is infinite or $-$ infinite.

Example. $f \in L^1[0,1]$, then (if $f \geq 0$, get a measure)

$$\nu(E) = \int_E f dx$$

(Positive sets): A set $A \in \mathcal{F}$ is a positive set if $\nu(E) \geq 0$ for every mble subset $E \subset A$.

(Negative sets): A set $A \in \mathcal{F}$ is negative if for every measurable subset $E \subset A$, $\nu(E) \leq 0$.

(Null): A set which is both positive and negative is a null set. Thus a set is null iff every measurable subset has measure zero.

Remark. Null sets are not the same as sets of measure zero.

Example. Take ν given above.

Our goal now is to prove that the space Ω can be written as the disjoint union of a positive set and a negative set. This is called the *Hahn–Decomposition*.

Lemma 1.1. *(i) Every mble subset of a positive set is positive.*

(ii) If $A_1, A_2, \ldots$ are positive then $A = \bigcup_{i=1}^{\infty} A_i$ is positive.

Proof. (i): Trivial.

Proof of (ii). : Let $A = \bigcup_{n=1}^{\infty} A_i$. A_i positive. Let $E \subset A$ be mble. Write

$$\begin{cases} E = \bigcup\limits_{n=1}^{\infty} E_i, \ E_i \cap E_j = 0, \ i \neq j \\ E_j = E \cap A_j \cap A_{j-1}^c \cap \ldots \cap A_1^c \subset A_j \end{cases} \tag{*}$$

$$\Rightarrow \nu(E_j) \geq 0$$

$$\nu(F) = \Sigma \nu(E_j) \geq 0$$

We show $(*)$: Let $x \in E_j \Rightarrow x \in E$ and $x \in E_j$ but $x \notin A_{j-1}, \ldots A_1$. $\therefore x \notin E_i$ if $j > i$. If $x \in E$, let $j =$ first j such that $x \in A_j$. Then $x \in E_j$, done. (Such a j exists become $E \subset A$).

Lemma 1.2. *Let E be measurable with $0 < \nu(E) < \infty$. Then there is mble set $A \subset E$. A positive such that $0 < \nu(A)$.*

Proof. If E is positive we are done. Let $n_1 =$ smallest positive number such that there is an $E_1 \subset E$ with

$$\nu(E_1) \leq -1/n_1.$$

Now, consider $E|E_1 \subset E$.

Again, if $E|E_1$ is positive with $\nu(E|E_1) > 0$ we are done. If not n_2 = smallest integer such that

1) $\exists E_2 \subset E|E_1$ with $\nu(E_2) < 1/n_2$.

Continue:

Let n_k = smallest positive integer such that

$$\exists E_k \subset E \Big| \bigcup_{j=1}^{k-1} E_j$$

with

$$\nu(E_k) < -\frac{1}{n_k}.$$

Let

$$A = E | \bigcup_{k=1}^{\infty} E_k.$$

Claim: A will do.

First: $\nu(A) > 0$. Why?

$$E = A \cup \bigcup_{k=1}^{\infty} E_k \text{ are disjoint.}$$

$$\nu(E) = \nu(A) + \sum_{k=1}^{\infty} \nu(E_k)$$

$$\Rightarrow \nu(A) > 0$$

since negative.

Now,

$$0 < \nu(E) < \infty \Rightarrow \sum_{k=1}^{\infty} \nu(E_k)$$

converges.

Problem 1: Prove that A is also positive.

Absolutely. $\therefore$

$$\sum_{k=1}^{\infty} \frac{1}{n_k} < -\sum \nu(E_k) < \infty.$$

Suppose A is not positive. Then A has a subset A_0 with $A_0) < -\varepsilon$ for some $\varepsilon > 0$. Now, since $\Sigma \frac{1}{n_k} < \infty,\ n_k \to \infty$.

Theorem 1.1 (Hahn–Decomposition)**.** *Let ν be a signed measure on $(\Omega, \mathcal{F})$. There is a positive set A and a negative set B with $A \cap B = \phi,\ A \cup B = X$.*

Proof. Assume ν does not take $+\infty$. Let $\lambda = \sup\{\nu(A) \colon A \in$ positive *sets*$\}$. $\phi \in$ Positive sets, sup ≥ 0. Let $A_n \in p$ s.t.

$$\lambda = \lim_{n \to \infty} \nu(A_n).$$

Set

$$A = \bigcup_{n=1}^{\infty} A_n.$$

A is positive. Also, $\lambda \geq \nu(A)$. Since

$$A \backslash A_n \subset A \Rightarrow \nu(A|A_n) \geq 0$$

and

$$\nu(A) = \nu(A_n) + \nu(A|A_n) \geq \nu(A_n).$$

Thus,

$$\nu(A) \geq \lambda \Rightarrow 0 \leq \nu(A) = \lambda < \infty.$$

$\therefore \quad 0 \leq \nu(A).$

Let $B = A^c$.

Claim B is negative.

Let $E \subset B$ and E positive.
We show $\nu(E) = 0$. This will do it.

For suppose $E \subset B,\ 0 < \nu(E) < \infty \Rightarrow E$ has a positive subset of positive measure by Lemma 1.2.

To show $\nu(E) = 0$, observe $E \cup A$ is positive

$$\begin{aligned} \therefore\ \lambda \geq \nu(E \cup A) &= \nu(E) + \nu(A) \\ &= \nu(E) + \lambda \Rightarrow \nu(E) = 0. \end{aligned}$$

Q.E.D.

Problem 1.b: Give an example to show that the Hahn decomposition is not unique.

Remark 1.1. The Hahn decomposition give two measures ν^+ and ν^- defined by

$$\begin{aligned} \nu^+(E) &= \nu(A \cap E) \\ \nu^-(E) &= -\nu(B \cap E). \end{aligned}$$

Notice that $\nu^+(B) = 0$ and $\nu^-(A) = 0$. Clearly $\nu(E) = \nu^+(E) - \nu^-(E)$.

Definition 1.2. Two measures ν_1 and ν_2 are mutually singular ($\nu_1 \perp \nu_2$) if there are two measurable subsets A and B with $A \cap B = \phi\ A \cup B = \Omega$ and $\nu_1(A) = \nu_2(B) = 0$. Notice that $\nu^+ \perp \nu^-$.

Theorem 1.2 (Jordan Decomposition)**.** *Let ν be a signed measure. These are two mutually singular measures ν^+ and ν^- such that*

$$\nu = \nu^+ - \nu^-.$$

This decomposition is unique.

Example. $f \in L^1[a,b]$

$$\nu(E) = \int_E f dx.$$

Then

$$\nu^+(E) = \int_E f^+ dx,\ \nu^-(E) = -\int_E f^- dx.$$

Definition 1.3. The measure ν is absolutely continuous with respect to μ, written $\nu << \mu$, if $\mu(A) = 0$ implies $\nu(A) = 0$.

Example. Let $f \geq 0$ be mble and set $\nu(A) = \int\limits_A f d\mu$.

$$\boxed{\nu << \mu.}$$

Theorem 1.3 (Radon–Nikodym Theorem)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be σ–finite measure spaces. Assume $\nu << \mu$. Then there is a nonnegative measurable function f such that*

$$\nu(E) = \int_E f d\mu.$$

The function f is unique a.e. $[\mu]$. We call f the Radon–Nikodym derivative of ν with respect to μ and write

$$f = \frac{d\nu}{d\mu}.$$

Remark 1.2. The space needs to be σ–finite.

Example. $(\Omega, \mathcal{F}, \mu) = ([0,1], \text{Borel}, \mu = \text{counting})$. Then

$$m << \mu, \quad m = \text{ Lebesgue}.$$

If

$$m(E) = \int_E f d\mu$$

$\Rightarrow f(x) = 0$ for all $x \in [0,1]$.

$$\therefore\ m = 0, \quad \text{Contra.}$$

3. Lemmas.

Lemma 1.3. *Suppose $\{B_\alpha\}_{\alpha\in D}$ is a collection of mble sets index by a countable set of real numbers D. Suppose $B_\alpha \subset B_\beta$ whenever $\alpha < \beta$. Then there is a mble function f such that $f(x) \le \alpha$ on B_α and $f(x) \ge \alpha$ on B_α^c.*

Proof. For $x \in \Omega$, set

$$\begin{aligned} f(x) &= \text{ first } \alpha \text{ such that } x \in B_\alpha \\ &= \inf\{\alpha \in D\colon x \in B_\alpha\}. \end{aligned}$$
$$\inf\{\phi\} = \infty.$$

- If $x \notin B_\alpha,\ x \notin B_\beta$ for any $\beta < \alpha$ and so, $f(x) \ge \alpha$.
- If $x \in B_\alpha$, then $f(x) \le \alpha$ provided we show f is mble. Q.E.D.

Claim: $\forall \lambda$ real

$$\{x\colon f(x) < \lambda\} = \bigcup_{\substack{\beta<\lambda \\ \beta\in D}} B_\beta.$$

If $f(x) < \lambda$, then $x \in D_\beta$ save $\beta < \lambda$. If $x \in B_\beta,\ \beta < \lambda \Rightarrow f(\alpha) < \lambda$. Q.E.D.

Lemma 1.4. *Suppose $\{B_\alpha\}_{\alpha\in D}$ as in Lemma 1.3 but this time $\alpha < \beta$ implies only $\mu\{D_\alpha \backslash B_\beta\} = 0$. Then there exists a mble function f on Ω such that $f(x) \le \alpha$ a.e. on B_α and $f(x) \ge \alpha$ a.e. on B_α^c.*

Lemma 1.5. *Suppose D is dense. Then the function in Lemma 1.3 is unique and the function in lemma 1.4 is unique μ a.e.*

Proof of Theorem 1.3. Assume $\mu(\Omega) = 1$. Let

$$\nu_\alpha = \nu - \alpha\mu, \quad \alpha \in \mathbb{Q}.$$

ν_α is a signed measure. Let $\{A_\alpha, B_\alpha\}$ be the Hahn–Decomp of ν_α. Notice:

$$\Omega = B_\alpha,\ B_\alpha = \phi,\ \text{if } \alpha \le 0. \tag{1}$$

$$B_\alpha | B_\beta = B_\alpha \cap (X|B_\beta) = B_\alpha \cap A_\beta. \tag{2}$$

Thus,

$$\nu_\alpha(B_\alpha|B_\beta) \leq 0 \tag{1}$$

$$\nu_\beta(B_\alpha|B_\beta) \geq 0 \tag{2}$$

or

$$\Leftrightarrow \nu(B_\alpha|B_\beta) - \alpha\mu(B_\alpha|B_\beta) \leq 0 \tag{1}$$

$$\nu(B_\alpha|B_\beta) - \beta\mu(B_\alpha|B_\beta) \geq 0. \tag{2}$$

Thus,

$$\beta\mu(B_\alpha|B_\beta) \leq \nu(B_\alpha|B_\beta) \leq \alpha\mu(B_\alpha|B_\beta).$$

Thus, if $\alpha < \beta$, we have

$$\mu(B_\alpha|B_\beta) = 0.$$

Thus, $\exists n$ mble f s.t. $\forall\ \alpha \in Q,\ f \geq \alpha$ a.e. on A_α and $f(x) \leq \alpha$ a.e. on B_α. Since $B_0 = \phi,\ f \geq 0$ a.e.

Let N be very large. Put

$$E_k = E \cap \left(\frac{B_{k+1}}{N}\middle| B_{k/N}\right),\ k = 0, 1, 2, \ldots$$

$$E_\infty = \Omega \Big| \bigcup_{k=0}^{\infty} B_{k/N}.$$

Then $E_0, E_1, \ldots, E_\infty$ are disjoint and

$$E = \bigcup_{k=0}^{\infty} E_k \cup E_\infty.$$

So,

$$\boxed{\nu(E) = \nu(E_\infty) + \sum_{k=0}^{\infty} \nu(E_k)}$$

on

$$E_k \subset \frac{B_{k+1}}{N}\Big| B_{k/N} = \frac{B_{k+1}}{N} \cap A_{k/N}.$$

We have,

$$k/N \le f(x) \le \frac{k+1}{N} \text{ a.e.}$$

and so,

$$\frac{k}{N}\mu(E_k) \le \int_{E_k} f(x)d\mu \le \frac{k+1}{N}\mu(E_k). \tag{1}$$

Also

$$E_k \subset A_{k/N} \Rightarrow \frac{k}{N}\mu(E_k) \le \nu(E_k) \tag{2}$$

and

$$E_k \subset \frac{B_{k+1}}{N} \Rightarrow \nu(E_k) \le \frac{k+1}{N}\mu(E_k). \tag{3}$$

Thus:

$$\begin{aligned}\nu(E_k) - \frac{1}{N}\mu(E_k) \le \frac{k}{N}\mu(E_k) &\le \int_{E_k} f(x)dx \\ &\le \frac{k}{N}\mu(E_k) + \frac{1}{N}\mu(E_k) \\ &\le \nu(E_k) + \frac{1}{N}\mu(E_k)\end{aligned}$$

on

$$E_\infty, \; f \equiv \infty \text{ a.e.}$$

If $\mu(E_\infty) > 0$, then $\nu(E_\infty) = 0$ since $(\nu - \alpha\mu)(E_\infty) \ge 0 \; \forall \; \alpha$. If $\mu(E_\infty) = 0 \Rightarrow \nu(E_\infty) = 0$. So, either way:

$$\nu(E_\infty) = \int_{E_\infty} f d\mu.$$

<u>Add</u>:

$$\nu(E) - \frac{1}{N}\mu(E) \le \int_E f d\mu \le \nu(E) + \frac{1}{N}\mu(E)$$

Since N is arbitrary, we are done.

Uniqueness: If $\nu(E) = \int\limits_E g d\mu, \; \forall \; E \in \mathcal{B}$

$$\Rightarrow \nu(E) - \alpha\mu(E) = \int_E (g - \alpha)d\mu \; \forall \; \alpha \; \forall \; E \subset A_\alpha.$$

Since

$$0 \leq \nu(E) - \alpha\mu(E) = \int_E (g - \alpha) d\mu.$$

We have:

$$g - \alpha \geq 0[\mu] \text{ a.e. on } A_\alpha$$

or

$$g \geq \alpha \text{ a.e. on } A_\alpha.$$

Similarly,

$$g \leq \alpha \text{ a.e. on } B_\alpha.$$

$\Rightarrow$

$$f = g \text{ a.e.}$$

Suppose μ is σ–finite: $\nu << \mu$. Let Ω_i be s.t. $\Omega_i \cap \Omega_j = \phi$, $\bigcup \Omega_i = \Omega$. $\mu(\Omega_i) < \infty$.

Put $\mu_i(E) = \mu(E \cap \Omega_i)$ and $\nu_i(E) = \nu(E \cap \Omega_i)$. Then $\nu_i << \mu_i$

$$\therefore \exists f_i \geq 0 \text{ s.t.}$$
$$\nu_i(E) = \int_E f_i d\mu_i$$

or

$$\nu(E \cap \Omega_i) = \int_{E \cap \Omega_i} f_i d\mu = \int_E d\Omega_i d\mu.$$

$\Rightarrow$ result.

Theorem 1.4 (The Lebesgue decomposition for measures)**.** *Let $(\Omega, \mathcal{F})$ be a measurable space and μ and ν σ–finite measures on $\mathcal{F}$. Then $\exists \nu_0 \perp \mu$ and $\nu_1 << \mu$ such that $\nu = \nu_0 + \nu_1$. The measure ν_0 and ν_1 are unique.*

$$(f \in BV \Rightarrow f = h + g.\ h \text{ singular } g \text{ a.c.}).$$

Proof. Let $\lambda = \mu + \nu$. λ is σ–finite.

$$\lambda(E) = 0 \Rightarrow \mu(E) = \nu(E) = 0.$$

R.N. $\Rightarrow$

$$\mu(E) = \int_E f d\lambda$$
$$\nu(E) = \int_E g d\lambda.$$

Let

$$A = \{f > 0\},\ B = \{f = 0\}$$
$$\Omega = A \cup B,\ A \cap B = \phi,\ \mu(B) = 0.$$

Let $\nu_0(E) = \nu(E \cap B)$. Then

$$\nu_0(A) = 0, \text{ so } \nu_0 \perp \mu$$

set

$$\nu_1(E) = \nu(E \cap A) = \int_{E \cap A} g d\lambda.$$

Clearly $\nu_1 + \nu_0 = \nu$ and it only remains to show that $\nu_1 << \mu$. Assume $\mu(E) = 0$. Then

$$\int_E f d\lambda = 0 \Rightarrow f \equiv 0 \text{ a.e. } [\lambda].\ (f \geq 0)$$

on E.

Since $f > 0$ on $E \cap A \Rightarrow \lambda(E \cap A) = 0$. Thus

$$\nu_1(E) = \int_{E \cap A} g d_\lambda = 0 \quad \text{Q.E.D.}$$

uniqueness. Problem.

You know: $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$, A, B. P and B indept. $P(A|B) = P(A)$. Now we work with probability measures. Let $(\Omega, \mathcal{F}_0, P)$ be a prob space, $\mathcal{F} \subset \mathcal{F}_0$ a σ–algebra and $X \in \sigma(\mathcal{F}_0)$ with $E|X| < \infty$.

Definition 1.4. The conditional expectation of X given $\mathcal{F}$, written as $E(X|\mathcal{F})$, is any random variable Y with the properties

(i) $Y \in \sigma(F)$

(ii) $\forall A \in \mathcal{F}$,

$$\int_A X dP = \int_A Y dP \text{ or } E(X;A) = E(Y;A).$$

<u>Existence</u>, <u>uniqueness</u>.

First, let us show that if Y has (i) and (ii), then $E|Y| \leq E|X|$.

With $A = \{Y > 0\} \in \mathcal{F}$, observe that

$$\int_A Y dp = \int_A X dp \leq \int_A |X| dp$$

and

$$\int_{A^c} -Y dp = \int_{A^c} -X dp \leq \int_{A^c} |X| dp.$$

So,

$$E|Y| \leq E|X|.$$

<u>Uniqueness</u>: If Y' also satisfies (i) and (ii), then

$$\begin{aligned} &\int_A Y dp = \int_A Y' dp \ \forall \ A \in \mathcal{F} \\ \Rightarrow &\int_A y dp = \int_A Y' dp \ \forall \ A \in \mathcal{F} \text{ or} \\ &\int_A (Y - Y') dp = 0 \ \forall \ A \in \mathcal{F}. \\ \Rightarrow &\ Y = Y \text{ a.e.} \end{aligned}$$

<u>Existence</u>:

Consider ν defined on $(\Omega, \mathcal{F})$ by

$$\nu(A) = \int_A X dp. \qquad A \in \mathcal{F}.$$

ν is a signed measure. $\nu << P$.

$\therefore\ \exists\ Y \in \sigma(\mathcal{F})$ s.t.

$$\nu(A) = \int_A Y dp \quad \forall\ A \in \mathcal{F}.$$

$\therefore\ \int\limits_A Y dp = \int\limits_A x dp. \quad \forall\ A \in \mathcal{F}.$

Example 1. Let A, B be fixed sets in $\mathcal{F}_0$. Let

$$\mathcal{F} = \sigma\{B\} = \{\phi, \Omega, B, B^c\}.$$

$E(1_A|\mathcal{F})$?

This is a function so that thwn we integrate over sets in $\mathcal{F}$, we get integral of 1_A over the sets.

i.e.

$$\begin{aligned}\int_B E(1_A|\mathcal{F})dP &= \int_B 1_A dP \\ &= P(A \cap B).\end{aligned}$$

$$E(1_A|\mathcal{F})P(B) = P(A \cap B)$$

or

$$E(1_A|\mathcal{F})1_B(B) = P(A \cap B)$$

or

$$P(A|B) = \boxed{E(1_A|\mathcal{F})1_B = \frac{P(A \cap B)}{P(B)}}.$$

<u>In general</u>. If X is a random variable and $\mathcal{F} = \sigma(\Omega_1, \Omega_2, \dots)$ where Ω_i are disjoint, then

$$\boxed{1_{\Omega_i} E(X|\mathcal{F}) = \frac{E(X;\Omega_i)}{P(\Omega_i)}}$$

or

$$\boxed{E(X|\mathcal{F}) = \sum_{i=1}^{\infty} \frac{E(X;\Omega_i)}{P(\Omega_i)}\ 1_{\Omega_i}(\omega).}$$

Notice that if $\mathcal{F}\{\phi, \Omega\}$.

Thus

$$E(X|\mathcal{F}) = E(X).$$

Properties:

(1) If $X \in \mathcal{F} \Rightarrow E(X|\mathcal{F}) = X$. $E(X) = E(E(X|\mathcal{F}))$.

(2) X is independent of $\mathcal{F}$, i.e. $(\sigma(X) \perp \mathcal{F})$

$$P(X \in B) \cap A) = P(X \in B)P(A).$$

$\Rightarrow E(X|\mathcal{F}) = E(X)$ (as in $P(A|B) = P(A)$).

Proof. To check this, (1) $E(X) \in \mathcal{F}$.

(ii) Let $A \in \mathcal{F}$. Thus

$$\begin{aligned}\int_A EX dP &= E(X)E(1_A) \\ &= E(X1_A) = \int_A X dp\end{aligned}$$

$\therefore\ E(X|\mathcal{F}) = EX.$

Theorem 1.5. *Suppose that X, Y and X_n are integrable.*

(i) If $X = a \Rightarrow E(X|\mathcal{F}) = a$

(ii) For constants a and b, $E(aX + bY) = aE(X|\mathcal{F}) + bE(Y|\mathcal{F})$.

(iii) If $x \le Y \Rightarrow E(X|\mathcal{F}) \le E(X|\mathcal{F})$. In particular, $|E(X|\mathcal{F})| \le E(|X||\mathcal{F})$.

(v) If $\lim_{n\to\infty} X_n = X$ and $|X_n| \le Y$, Y integrable, then $E(X_n|\mathcal{F}) \to E(X|\mathcal{F})$ a.s.

(vi) Monotone and Fatou's hold the same way.

Proof. (i) Done above since clearly $a \in \mathcal{F}$.

(ii) $\int\limits_A aE(x|\mathcal{F})dp + \int\limits_A bE(Y|\mathcal{F})dP$

$$\begin{aligned}
&\int_A (aE(x|\mathcal{F}) + bE(Y|I))dp = \\
&= a\int_A E(X|\mathcal{F})dP + b\int_A E(Y|\mathcal{F})dP \\
&= a\int_A XdP + b\int_A Ydp = \int_A (aX+bY)dp = \int_A E(aX+bY|\mathcal{F})dP
\end{aligned}$$

(iii)

$$\begin{aligned}
\int_A E(X|\mathcal{F})dP = \int_A XdP \leq \int_A YdP \\
= \int_A E(Y|\mathcal{F})dP. \quad \forall\ A \in \mathcal{F}
\end{aligned}$$

$\therefore E(X|\mathcal{F}) \leq E(Y|\mathcal{F})$. a.s.

(iv) Let $Z_n = \sup\limits_{k\geq n} |X_k - X|$. Then $Z_n \downarrow 0$ a.s.

$$\begin{aligned}
|E(X_n|\mathcal{F}) - E(X|\mathcal{F})| &\leq E(|X_n - Y||\mathcal{F}_n) \\
&\leq E[Z_n|\mathcal{F}].
\end{aligned}$$

Need to show $E[Z_n|\mathcal{F}) \downarrow 0$ with pub. 1. By (iii), $E(Z_n|\mathcal{F}_n) \downarrow$ decreasing. So, let $Z =$ limit. Need to show $Z \equiv 0$.

We have $Z_n \leq 2Y$.

$$\begin{aligned}
\therefore\ E(Z) &= \int_\Omega E(Z|\mathcal{F})dp \quad (E(Z| = E(E(Z|\mathcal{F})) \\
&= E(E(Z|\mathcal{F})) \leq E(E(Z_n|\mathcal{F})) = E(Z_n) \to 0 \text{ by D.C.T.}
\end{aligned}$$

Theorem 1.6 (Jensen Inequality)**.** *If φ is convex $E|X|$ and $E|\varphi(X)| < \infty$, then*

$$\varphi(E(X|\mathcal{F})) \leq E(\varphi(X)|\mathcal{F}).$$

Proof.

$\varphi(x_0) + A(x_0)(x - x_0) \le \varphi(x)$.
$x_0 = E(X|\mathcal{F})$ $x = X$.

$$\varphi(E(X|\mathcal{F})) + A(E(X|\mathcal{F}))(X - E(X|\mathcal{F})) \le \varphi(X).$$

Note expectation of both sides.

$$E[\varphi(E(X|\mathcal{F})) + A(E(X|\mathcal{F}))(X - E(X|\mathcal{F})|\mathcal{F}) \le E(\varphi(X|\mathcal{F})$$
$$\Rightarrow \varphi(E(X|\mathcal{F})) + A(E(X|\mathcal{F}))[E(X|\mathcal{F}) - E(X|\mathcal{F})] \le E(\varphi(X|\mathcal{F})).$$

Corollary.

$$|E(X|\mathcal{F})|^p \le E(|X|^p|\mathcal{F}) \text{ for } 1 \le p < \infty$$
$$\exp(E(X|\mathcal{F})) \le E(e^X|\mathcal{F}).$$

Theorem 1.7. *(1) If $\mathcal{F}_1 \subset \mathcal{F}_2$ then*

(a) $E(E(X|\mathcal{F}_1)|\mathcal{F}_2)) = E(X|\mathcal{F}_1)$

(b) $E(E(X|\mathcal{F}_2)|\mathcal{F}_1)) = E(X|\mathcal{F}_1)$. *(The smallest field always wins).*

(2) If $X \in \sigma(\mathcal{F})$, $E|Y|$, $E|XY| < \infty$

$$\Rightarrow E(XY|\mathcal{F}) = XE(Y|\mathcal{F})$$

(measurable functions act like constants). ($Y = 1$, done before).

Proof. (a) $E(X|\mathcal{F}_1) \in (\mathcal{F}_1) \subset (\mathcal{F}_2)$.
$\therefore$ Done.

(b) $E(X|\mathcal{F}_1) \in \mathcal{F}_1$. So $A \in \mathcal{F}_1 \subset \mathcal{F}_2$ we have

$$\int_A E(X|\mathcal{F}_1)dp = \int_A (X)dp$$
$$= \int_A E(X|\mathcal{F}_2)dp = \int_A E(E(x|\mathcal{F}_2)|\mathcal{F}_2)dP.$$

Durrett: p 220: #1.1

p. 222: #1.2

p.225: #1.3

p. 227: 1.6

p. 228: 1.8.

Let

$$L^2(\mathcal{F}_0) = \{X \in \mathcal{F}_0 \colon EX^2 N\infty\}$$

and

$$L^2(\mathcal{F}_1) = \{Y \in \mathcal{F}_1 \colon EY^2 < \infty\}.$$

Then $L^2(\mathcal{F}_1)$ is a closed subspace of $L^2(\mathcal{F}_0)$. In fact, with $\langle X_1, X_2\rangle = E(X_1 \cdot X_2)$, $L^2(\mathcal{F}_0)$ and $L^2(\mathcal{F}_1)$ are Hilbert spaces. $L^2(\mathcal{F}_1)$ is closed subspace in $L^2(\mathcal{F}_0)$. Given any $X \in L^2(\mathcal{F}_0)$, $\exists\, Y \in L^2(\mathcal{F}_1)$ such that

$$\text{dist }(X, L^2(\mathcal{F}_1)) = E(x-y)^2$$

Theorem 1.8. *Suppose $Ex^2 < \infty$, then*

$$\inf_{X \in L^2(\mathcal{F}_1)} E(|X-Y|^2) = E(|X-(EX|\mathcal{F}_1))^2.$$

Proof. Need to show

$$E(|X-Y|^2) \geq E|X - E(X|\mathcal{F}_1))^2$$

for any $y \in L(\mathcal{F}_1)$. Let $y \in L^2(\mathcal{F}_1)$ out set. Set

$$Z = Y - E(X|\mathcal{F}) \in L^2(\mathcal{F}_1),$$
$$Y = Z + E(X|\mathcal{F}).$$

Now, since

$$E(ZE(X|\mathcal{F})) = E(ZX|\mathcal{F})) = E(ZX)$$

we see that

$$
\begin{aligned}
&E(ZE(X|\mathcal{F})) - E(ZX) = 0.\\
&E(X-Y)^2 = E\{X - Z - E(X|\mathcal{F})\}^2\\
&= E(X - E(X|\mathcal{F}))^2 + E(Z^2)\\
&\qquad - 2E((X - E(X|\mathcal{F}))Z)\\
&\geq E(X - E(X|\mathcal{F}))^2\\
&\qquad - 2E(XZ) + 2E(ZE(X|\mathcal{F}))\\
&= E(X - E(X|\mathcal{F}))^2. \qquad \text{Q.E.D.}
\end{aligned}
$$

By the way: If X and Y are two r.v. we define

$$E(X|Y) = E(x|\sigma(Y)).$$

Recall conditional expectation for ??

Suppose X and Y have joint density $f(x,y)$

$$P((X,Y)\in\mathcal{B}) = \int_B f(x,y)dxdy, \quad \mathcal{B}\subset\mathbb{R}^2.$$

And suppose $\int\limits_{\mathbb{R}} f(x,y)dx > 0$ $\forall y$. We claim that in this case, if $E|g(X)| < \infty$, then

$$E(g(X)|Y) = h(Y)$$

and

$$h(y) = \frac{\int_{\mathbb{R}} g(x)f(x,y)dy}{\int_{\mathbb{R}}(f(x,y)dx}.$$

Treat the "given" density as if the second probability

$$P(X=x|Y=y) = \frac{P(X=x,\ Y=y)}{P(Y=y)}$$

$$= \frac{f(x,y)}{\int_{\mathbb{R}} f(x,y)dx}.$$

Now: <u>Integrale</u>:

$$E(g(X)|Y = y) = \int g(x)P(X = x)Y = y))dy.$$

<u>To verify</u>: (i) clearly $h(Y) \in \sigma(Y)$. For (ii): let

$$A = \{Y \in B\} \text{ for } B \in \mathcal{B}(\mathbb{R}).$$

Then <u>need to show</u>

$$\boxed{E(h(Y); A) = E(g(X); A).}$$

L.H.S., $E(h(Y)1_{\mathbb{R}}(X) \cdot A)$

$$\begin{aligned}
&= \int_B \int_{\mathbb{R}} h(y)f(x,y)dxdy \\
&\int_B \int_{\mathbb{R}} \frac{\left(\int\limits_{\mathbb{R}} g(3)f(3,y)d_3\right)}{\int\limits_{\mathbb{R}} f(x,y)dx} f(x,y)dxdy \\
&= \int_B \int_{\mathbb{R}} g(3)f(3,y)d_3 d_y \\
&= E(g(X)1_B(Y)) = E(g(X); A).
\end{aligned}$$

(If $\int f(x,y)dy = 0$, define h by $h(y)\int f(x,y)dx = \int g(x)f(x,y)dy$ i.e. h can be anything where $\int f(x,y)\ dy = 0$).