Notes on Probability Theory

Kai-Jyun Wang*

Fall 2025

The notes are based on the lecture of Prof. David Anderson at University of Wisconsin-Madison in fall 2025. The course structure mainly follows Durrett.

*National Taiwan University, Department of Economics.

Contents

1. Probability Space

1.1. Probability Space

Definition 1.1
*Let Ω be a set. A collection of subsets $\mathcal{F}$ forms a σ-**algebra** if*

(a) $\varnothing \in \mathcal{F}$.

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

(c) *If* $A_i \in \mathcal{F}$ *are countably many sets,* $\cup_i A_i \in \mathcal{F}$.

The dual $(\Omega, \mathcal{F})$ *is called a **measurable space** and the sets falling in* $\mathcal{F}$ *are said to be **measurable**.*

Definition 1.2
Let $(\Omega, \mathcal{F})$ *be a measurable space. A set function* $\mu : \mathcal{F} \to [0, \infty]$ *is a **measure** if*

(a) $\mu(\varnothing) = 0$.

(b) *For countably many disjoint* $A_i \in \mathcal{F}$, $\mu(\cup_i A_i) = \sum_i \mu(A_i)$.

The triple $(\Omega, \mathcal{F}, \mu)$ *is called a **measure space**.*

Definition 1.3
*A **probability space** is a measure space* $(\Omega, \mathcal{F}, \mathrm{P})$ *such that* $\mathrm{P}(\Omega) = 1$.

Lemma 1.4
Let $\mathcal{S}$ *be a collection of sets. Then there exists the smallest* σ*-algebra containing* $\mathcal{S}$.

Proof. Let $\mathcal{F}$ be the intersection of all σ-algebra containing $\mathcal{S}$. $\mathcal{F}$ is non-empty since the power set is a σ-algebra containing $\mathcal{S}$. Now it is clear that $\varnothing \in \mathcal{F}$ since $\varnothing \in \mathcal{A}$ for every σ-algebra $\mathcal{A}$ contiaining $\mathcal{S}$. If $A \in \mathcal{F}$, $A \in \mathcal{A}$ for all $\mathcal{A}$ containing $\mathcal{S}$ and $A^c \in \mathcal{A}$ for all $\mathcal{A}$. Thus $A^c \in \mathcal{F}$. Finally, if $A_i \in \mathcal{F}$ are countably many sets, then each A_i lies in every $\mathcal{A}$ containing $\mathcal{S}$; so does $\cup_i A_i$ and thus $\cup_i A_i \in \mathcal{F}$. The minimality follows by the construction of $\mathcal{F}$. ■

Definition 1.5
For any collection of sets $\mathcal{S}$*, the smallest* σ*-algebra is denoted as* $\sigma(\mathcal{S})$.

Theorem 1.6
Let $(\Omega, \mathcal{F}, \mathrm{P})$ *be a probability space. Then*

(a) *If* $A, B \in \mathcal{F}$ *and* $A \subset B$*, then* $\mathrm{P}(A) \leq \mathrm{P}(B)$.

(b) *For countably many* $A_i \in \mathcal{F}$, $\mathrm{P}(\cup_i A_i) \leq \sum_i \mathrm{P}(A_i)$.

(c) *If* $A_i \nearrow A$, $\mathrm{P}(A_i) \to \mathrm{P}(A)$.

(d) *If* $A_i \searrow A$, $\mathrm{P}(A_i) \to \mathrm{P}(A)$.

Proof. (a) and (b) are clear. For (c), write $E_i = A_i - A_{i-1}$ and $A_0 = \varnothing$. Then since E_i are disjoint and $A_n = \cup_{i=1}^n E_i$,

$$\mathrm{P}(A_n) = \mathrm{P}\big(\cup_{i=1}^n E_i\big) = \sum_{i=1}^n \mathrm{P}(E_i) \to \sum_i \mathrm{P}(E_i) = \mathrm{P}(\cup_i E_i) = \mathrm{P}(A)$$

as $n \to \infty$.

For (d), note that $A_i^c \nearrow A^c$. Thus $1 - \mathrm{P}(A_i) = \mathrm{P}(A_i^c) \to \mathrm{P}(A^c) = 1 - \mathrm{P}(A)$. Thus $\mathrm{P}(A_i) \to \mathrm{P}(A)$. ∎

Definition 1.7
*The **Borel σ-algebra** is the σ-algebra generated by all open sets.*

Definition 1.8
Let P *be a probability measure on* $(\mathbb{R}, \mathcal{B})$. *The **distribution function** F is defined as*

$$F(x) = \mathrm{P}((-\infty, x])$$

for $x \in \mathbb{R}$.

Proposition 1.9
The distribution function in $(\mathbb{R}, \mathcal{B})$ *satisfies that*

(a) $F(x) \leq F(y)$ *for all* $x \leq y$.

(b) $F(x) \to F(y)$ *as* $x \to y^+$.

(c) $F(-\infty) = 0$ *and* $F(\infty) = 1$.

Proof. For (a), note that $(-\infty, x] \subset (-\infty, y]$ and

$$F(x) = \mathrm{P}((-\infty, x]) \leq \mathrm{P}((-\infty, y]) = F(y).$$

For (b), notice that for $x_n \to y^+$, $(-\infty, x_n] \searrow (-\infty, y]$. Hence

$$F(x_n) = \mathrm{P}((-\infty, x_n]) \to \mathrm{P}((-\infty, y]) = F(y).$$

Similarly, taking $x_n \to \pm\infty$ gives (c). ∎

Definition 1.10
*A collection $\mathcal{S}$ of sets is called an **algebra** if*

(a) $\varnothing \in \mathcal{S}$.

(b) If $A \in \mathcal{S}$, *then* $A^c \in \mathcal{S}$.

(c) If $A, B \in \mathcal{S}$, *then* $A \cup B \in \mathcal{S}$.

Remark
An algebra is closed under finite unions. It is also clear that a σ-algebra is an algebra, while the converse is not true. An example is the collection of all finite unions of intervals in $\mathbb{R}$.

Definition 1.11

A collection $\mathcal{S}$ of sets is called a ***semi-algebra*** *if*

(a) If $A, B \in \mathcal{S}$, then $A \cap B \in \mathcal{S}$.

(b) If $A \in \mathcal{S}$, then A^c can be written as a finite disjoint union of sets in $\mathcal{S}$.

Remark

A semi-algebra must contain $\varnothing$ since for any $A \in \mathcal{S}$, $A^c = \cup_i A_i$, where $A_i \in \mathcal{S}$ are disjoint. Then $A \cap A_1 = \varnothing \in \mathcal{S}$.

Remark

An example of being a semi-algebra but not an algebra is the collection of all intervals of the form $(a_i, b_i]$ for $-\infty \le a_i < b_i \le \infty$ with the empty set.

Lemma 1.12

If $\mathcal{S}$ is a semi-algebra, then $\overline{\mathcal{S}} = \{$finite disjoint unions of sets in $\mathcal{S}\}$ forms an algebra.

Proof. It has been shown that $\varnothing \in \mathcal{S}$. For $A, B \in \overline{\mathcal{S}}$, write $A = \cup_{i=1}^n A_i$ and $B = \cup_{j=1}^m B_j$ for disjoint $A_i, B_j \in \mathcal{S}$, respectively. Then $A \cap B = \cup_{i,j}(A_i \cap B_j) \in \overline{\mathcal{S}}$. Thus $\overline{\mathcal{S}}$ is closed under intersection. Now if $A \in \overline{\mathcal{S}}$, $A = \cup_{i=1}^n A_i$ for disjoint $A_i \in \mathcal{S}$. Then $A^c = \cap_{i=1}^n A_i^c$. By the definition of semi-algebra, A_i^c can be written as a finite disjoint union of sets in $\mathcal{S}$ and thus $A_i^c \in \overline{\mathcal{S}}$. Since $\overline{\mathcal{S}}$ is closed under finite intersection, $A^c = \cap_{i=1}^n A_i^c \in \overline{\mathcal{S}}$. Finally, for $A, B \in \overline{\mathcal{S}}$, $A \cup B = (A^c \cap B^c)^c \in \overline{\mathcal{S}}$. We conclude that $\overline{\mathcal{S}}$ is indeed an algebra. ∎

Definition 1.13

Suppose $\mathcal{S}$ is a semi-algebra. $\overline{\mathcal{S}} = \{$finite disjoint unions of sets in $\mathcal{S}\}$ is called the ***algebra generated by*** *$\mathcal{S}$.*

Definition 1.14

Let $\mathcal{S}$ be an algebra. A set function $\mu_0 : \mathcal{S} \to [0, \infty]$ is called a ***premeasure*** *if*

(a) $\mu_0(\varnothing) = 0$.

(b) For countable disjoint $A_i \in \mathcal{S}$ such that $\cup_i A_i \in \mathcal{S}$,

$$\mu_0(\cup_i A_i) = \sum_i \mu_0(A_i).$$

Theorem 1.15

Let ν be a set function on a semi-algebra $\mathcal{S}$ such that $\nu(\varnothing) = 0$. Suppose that

(a) if $A \in \mathcal{S}$ and $A = \cup_{i=1}^n A_i$ for disjoint $A_i \in \mathcal{S}$, then $\nu(A) = \sum_{i=1}^n \nu(A_i)$;

(b) if $A_i \in \mathcal{S}$ are countably many sets and $A = \cup_i A_i \in \mathcal{S}$, then $\nu(A) \le \sum_i \nu(A_i)$.

Then ν can be extended to a unique premeasure μ_0 on the algebra generated by $\mathcal{S}$.

Proof. We first show the existence. From lemma 1.12 we know that $\mathcal{S}$ generates an algebra $\mathcal{A} = \{$finite disjoint union of sets in $\mathcal{S}\}$. Define our candidate μ_0 by $\mu_0(A) = \sum_i \nu(A_i)$ for

$A = \cup_i A_i$ where $A_i \in \mathcal{S}$ are disjoint. To see that μ_0 is well-defined, suppose $A = \cup_i B_i$ for disjoint $B_i \in \mathcal{S}$. Observe that

$$A_i = \cup_j (A_i \cap B_j) \quad \text{and} \quad B_j = \cup_i (A_i \cap B_j)$$

are finite disjoint unions. Then

$$\sum_i \nu(A_i) = \sum_i \sum_j \nu(A_i \cap B_j) = \sum_j \sum_i \nu(A_i \cap B_j) = \sum_j \nu(B_j)$$

by (a). Thus μ_0 is well-defined.

Now we check that μ_0 is a premeasure. Clearly $\mu_0(\varnothing) = 0$. For finitely many disjoint $A_i \in \mathcal{A}$ such that $\cup_i A_i \in \mathcal{A}$, we can write $A_i = \cup_j B_{ij}$ for disjoint $B_{ij} \in \mathcal{S}$. Then (a) implies that

$$\mu_0(\cup_i A_i) = \mu_0(\cup_{i,j} B_{ij}) = \sum_{i,j} \nu(B_{ij}) = \sum_i \sum_j \mu_0(B_{ij}) = \sum_i \mu_0(A_i).$$

Next, for countably many disjoint $A_i \in \mathcal{A}$ such that $A = \cup_i A_i \in \mathcal{A}$, write $A_i = \cup_j B_{ij}$, where $B_{ij} \in \mathcal{S}$ are finite disjoint for each i. Then $\mu_0(A_i) = \sum_j \nu(B_{ij})$ and

$$\sum_i \mu_0(A_i) = \sum_i \sum_j \nu(B_{ij}).$$

Without loss of generality, we may choose A_i to be those in $\mathcal{S}$ since otherwise we can replace A_i by B_{ij}. We assume that $A_i \in \mathcal{S}$ from now on. Since $A \in \mathcal{A}$, $A = \cup_i C_i$ for finite disjoint $C_i \in \mathcal{S}$. $C_i = \cup_j (C_i \cap A_j)$. Thus (b) gives that

$$\nu(C_i) \le \sum_j \nu(C_i \cap A_j).$$

Then

$$\mu_0(A) = \sum_i \nu(C_i) \le \sum_i \sum_j \nu(C_i \cap A_j) = \sum_j \sum_i \nu(C_i \cap A_j) = \sum_j \nu(A_j) = \sum_j \mu_0(A_j).$$

For the opposite inequality, set $B_n = \cup_{i=1}^n A_i$ and $C_n = A - B_n$. Since $\mathcal{A}$ is an algebra, $C_n \in \mathcal{A}$ and the finite additivity shows that

$$\mu_0(A) = \sum_{i=1}^n \mu_0(A_i) + \mu_0(C_n) \ge \sum_{i=1}^n \mu_0(A_i).$$

Taking $n \to \infty$ gives the desired inequality and thus μ_0 is σ-additive on $\mathcal{A}$.

Finally, if μ_1 is another premeasure on $\mathcal{A}$ extending ν, then for $A = \cup_i A_i$ for disjoint $A_i \in \mathcal{S}$,

$$\mu_1(A) = \sum_i \nu(A_i) = \mu_0(A).$$

■

Definition 1.16
*A collection of sets $\mathcal{P}$ is called a π-**system** if $A, B \in \mathcal{P}$ implies $A \cap B \in \mathcal{P}$.*

Definition 1.17
*A collection of sets $\mathcal{L}$ is called a λ-**system** if*

(a) $\Omega \in \mathcal{L}$.

(b) If $A, B \in \mathcal{L}$ and $A \subset B$, then $B - A \in \mathcal{L}$.

(c) If $A_i \in \mathcal{L}$ and $A_i \nearrow A$, then $A \in \mathcal{L}$.

Theorem 1.18 (Sierpiński-Dynkin π-λ)
If $\mathcal{P}$ is a π-system and $\mathcal{L}$ is a λ-system containing $\mathcal{P}$, then $\sigma(\mathcal{P}) \subset \mathcal{L}$.

Proof. First we show that a collection $\mathcal{S}$ is a σ-algebra if and only if it is both a π-system and a λ-system. Suppose first that $\mathcal{S}$ is a π-system and a λ-system. $\varnothing = \Omega - \Omega \in \mathcal{S}$. If $A \in \mathcal{S}$, then $A^c = \Omega - A \in \mathcal{S}$. For $A, B \in \mathcal{S}$, $A \cup B = (A^c \cap B^c)^c \in \mathcal{S}$ since we have shown that $\mathcal{S}$ is closed under complement and intersection by being a π-system. Thus $\mathcal{S}$ is also closed under finite unions. If $A_i \in \mathcal{S}$ are countably many sets, let $B_n = \cup_{i=1}^n A_i \in \mathcal{S}$. Then $B_n \nearrow \cup_i A_i$ and thus $\cup_i A_i \in \mathcal{S}$.

Conversely, if $\mathcal{S}$ is a σ-algebra, then for $A, B \in \mathcal{S}$, $A \cap B = (A^c \cup B^c)^c \in \mathcal{S}$. Thus $\mathcal{S}$ is a π-system. If $A, B \in \mathcal{S}$ and $A \subset B$, then $B - A = B \cap A^c \in \mathcal{S}$. Finally, if $A_i \in \mathcal{S}$ and $A_i \nearrow A$, then $A = \cup_i (A_i - A_{i-1}) \in \mathcal{S}$ with $A_0 = \varnothing$. Thus $\mathcal{S}$ is a λ-system.

Now set $\mathcal{L}$ to be the smallest λ-system containing $\mathcal{P}$. It suffices to show that $\mathcal{L}$ is also a π-system and thus by the above conclusion, $\mathcal{L}$ is a σ-algebra containing $\mathcal{P}$; hence $\sigma(\mathcal{P}) \subset \mathcal{L}$.

To show that $\mathcal{L}$ is a π-system, let $A, B \in \mathcal{L}$. If $A, B \in \mathcal{P}$, $A \cap B \in \mathcal{P} \subset \mathcal{L}$. To extend the result for general $A, B \in \mathcal{L}$, we first fix $B \in \mathcal{P}$ and define

$$\mathcal{L}_B = \{A \mid A \cap B \in \mathcal{L}\}.$$

We claim that $\mathcal{L}_B$ is a λ-system containing $\mathcal{P}$. For $A \in \mathcal{P}$, $A \cap B \in \mathcal{L}$. Thus $\mathcal{P} \subset \mathcal{L}_B$. Clearly $\Omega \in \mathcal{L}_B$. If $E, F \in \mathcal{L}_B$ and $E \subset F$, then

$$(F - E) \cap B = (F \cap B) - (E \cap B) \in \mathcal{L}.$$

Thus $F - E \in \mathcal{L}_B$. Finally, if $E_i \in \mathcal{L}_B$ and $E_i \nearrow E$, then

$$E \cap B = \cup_i (E_i \cap B) \in \mathcal{L}.$$

Hence $E \in \mathcal{L}_B$ and we conclude that $\mathcal{L}_B$ is a λ-system. Since it is a λ-system containing $\mathcal{P}$, it also contains the smallest λ-system $\mathcal{L}$ with the intersection property. Thus $A \cap B \in \mathcal{L}$ whenever $A \in \mathcal{L}$ and $B \in \mathcal{P}$.

Next, fix $A \in \mathcal{L}$ and define $\mathcal{L}_A = \{B \mid A \cap B \in \mathcal{L}\}$. Clearly $\mathcal{L}_A$ contains $\mathcal{L}$ and $\Omega \in \mathcal{L}_A$. If $E, F \in \mathcal{L}_A$ and $E \subset F$, then

$$(F - E) \cap A = (F \cap A) - (E \cap A) \in \mathcal{L}.$$

Thus $F - E \in \mathcal{L}_A$. Finally, if $E_i \in \mathcal{L}_A$ and $E_i \nearrow E$, then

$$E \cap A = \cup_i (E_i \cap A) \in \mathcal{L}.$$

Hence $E \in \mathcal{L}_A$ and we conclude that $\mathcal{L}_A$ is a λ-system. Since it contains $\mathcal{L}$, $A, B \in \mathcal{L}$ implies $A \cap B \in \mathcal{L}$; in other words, $\mathcal{L}$ is a π-system and the proof is complete. ∎

Corollary 1.19
Let μ and ν be two probability measures agreeing on a π-system $\mathcal{P}$, i.e., $\mu(A) = \nu(A)$ for all $A \in \mathcal{P}$. Then $\mu(A) = \nu(A)$ for all $A \in \sigma(\mathcal{P})$.

Proof. Put

$$\mathcal{L} = \{A \mid \mu(A) = \nu(A)\}.$$

We claim that $\mathcal{L}$ is a λ-system containing $\mathcal{P}$. It is clear that by our assumption, $\mathcal{P} \subset \mathcal{L}$ and $\Omega \in \mathcal{L}$. If $A, B \in \mathcal{L}$ and $A \subset B$, then

$$\mu(B - A) = \mu(B) - \mu(A) = \nu(B) - \nu(A) = \nu(B - A).$$

Thus $B - A \in \mathcal{L}$. Finally, if $A_i \in \mathcal{L}$ and $A_i \nearrow A$, then

$$\mu(A) = \lim_{i \to \infty} \mu(A_i) = \lim_{i \to \infty} \nu(A_i) = \nu(A).$$

Hence $A \in \mathcal{L}$ and we conclude that $\mathcal{L}$ is a λ-system. By the Sierpiński-Dynkin π-λ theorem, $\sigma(\mathcal{P}) \subset \mathcal{L}$; in other words, μ and ν agree on $\sigma(\mathcal{P})$. ∎

Definition 1.20
*A measure μ on a measurable space $(\Omega, \mathcal{F})$ is called **σ-finite** if there exists countable $A_i \in \mathcal{F}$ such that $\cup_i A_i = \Omega$ and $\mu(A_i) < \infty$.*

Definition 1.21
*A set function $\mu^* : 2^\Omega \to [0, \infty]$ is called an **outer measure** if*

(a) $\mu^*(\varnothing) = 0$.

(b) If $A \subset B$, then $\mu^*(A) \le \mu^*(B)$.

(c) For countably many $A_i \subset \Omega$, $\mu^*(\cup_i A_i) \le \sum_i \mu^*(A_i)$.

Definition 1.22
*Let μ^* be an outer measure. A set $A \subset \Omega$ is said to be **Carathéodory measurable** or **μ^*-***

measurable *if for all* $E \subset \Omega$,

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c).$$

Lemma 1.23
Let μ^* be an outer measure on Ω. Then the collection of all μ^*-measurable sets forms a σ-algebra $\mathcal{F}$ and $\mu^*|_{\mathcal{F}}$ is a measure.

Proof. Put

$$\mathcal{F} = \{A \subset \Omega \mid \mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c) \text{ for all } E \subset \Omega\}.$$

We first show that $\mathcal{F}$ is a σ-algebra. Clearly $\varnothing \in \mathcal{F}$ and if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$. For $A, B \in \mathcal{F}$, let $C = A \cup B$. The property of outer measure gives that $\mu^*(E) \leq \mu^*(E \cap C) + \mu^*(E \cap C^c)$. To see the opposite inequality, note that $C = A \cup (B \cap A^c)$ and

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

Hence $C \in \mathcal{F}$ and $\mathcal{F}$ is closed under finite unions. For countable disjoint $A_i \in \mathcal{F}$ with $A = \cup_i A_i$, let $B_n = \cup_{i=1}^n A_i \in \mathcal{F}$. Then

$$\mu^*(E \cap A) = \mu^*(E \cap B_n) + \mu^*(E \cap B_n^c) \geq \mu^*(E \cap B_n) = \sum_{i=1}^n \mu^*(E \cap A_i).$$

Taking $n \to \infty$ gives that

$$\mu^*(E \cap A) \geq \sum_i \mu^*(E \cap A_i) \geq \mu^*(E \cap A)$$

by the σ-subadditivity of outer measure. Hence $\mu^*(E \cap A) = \sum_i \mu^*(E \cap A_i)$. Note also that $E \cap A^c \subset E \cap B_n^c$ so $\mu^*(E \cap A^c) \leq \mu^*(E \cap B_n^c)$. Thus

$$\mu^*(E) = \mu^*(E \cap B_n) + \mu^*(E \cap B_n^c) \geq \sum_{i=1}^n \mu^*(E \cap A_i) + \mu^*(E \cap A^c) \to \mu^*(E \cap A) + \mu^*(E \cap A^c) \geq \mu^*(E)$$

by the σ-subadditivity of outer measure. We conclude that $\mathcal{F}$ is a σ-algebra.

Finally, denote $\mu^*|_{\mathcal{F}}$ by μ. Clearly $\mu(\varnothing) = 0$. For countably many disjoint $A_i \in \mathcal{F}$ such that $A = \cup_i A_i \in \mathcal{F}$, let $B_n = \cup_{i=1}^n A_i \in \mathcal{F}$. Then

$$\mu(A) = \mu(B_n) + \mu(A \cap B_n^c) \geq \mu(B_n) = \sum_{i=1}^n \mu(A_i) \to \sum_i \mu(A_i) \geq \mu(A).$$

Hence $\mu(A) = \sum_i \mu(A_i)$ and μ is a measure on $\mathcal{F}$. ■

Theorem 1.24 (Carathéodory Extension)
Let ν be a finitely additive, σ-subadditive set function on a semi-algebra $\mathcal{S}$ such that $\nu(\varnothing) = 0$. Then ν can be extended to a measure on $\sigma(\mathcal{S})$.

Proof. By theorem 1.15, ν can be extended to a premeasure μ_0 on the algebra $\mathcal{A}$ generated by $\mathcal{S}$. Define the outer measure by

$$\mu^*(A) = \inf \left\{ \sum_i \mu_0(E_i) \,\middle|\, A \subset \cup_i E_i, E_i \in \mathcal{A} \right\}$$

for all $A \subset \Omega$ with the convention that $\inf \varnothing = \infty$. We check that μ^* is indeed an outer measure. Clearly $\mu^*(\varnothing) = 0$. If $A \subset B$, then any cover of B by sets in $\mathcal{A}$ is also a cover of A and hence $\mu^*(A) \leq \mu^*(B)$. For countably many $A_i \subset \Omega$, we can find $\{E_{ij}\}_j$ covering A_i such that

$$\sum_j \mu_0(E_{ij}) \leq \mu^*(A_i) + 2^{-i}\epsilon$$

for some $\epsilon > 0$. Then $\cup_{i,j} E_{ij}$ covers $\cup_i A_i$ and

$$\mu^*(\cup_i A_i) \leq \sum_i \sum_j \mu_0(E_{ij}) \leq \sum_i \mu^*(A_i) + \epsilon.$$

Since ϵ is arbitrary, we conclude that $\mu^*(\cup_i A_i) \leq \sum_i \mu^*(A_i)$ and μ^* is indeed an outer measure.

It follows from lemma 1.23 that the collection of all μ^*-measurable sets forms a σ-algebra $\mathcal{F}$ and μ^* restricted on $\mathcal{F}$ is a measure. It is clear that $\mathcal{A} \subset \mathcal{F}$ and $\sigma(\mathcal{S}) \subset \sigma(\mathcal{A}) \subset \mathcal{F}$ and $\mu = \mu^*|_{\sigma(\mathcal{S})}$ is also a measure. Finally, for $A, A_i \in \mathcal{S}$ where A_i covers A,

$$\mu(A) = \mu^*(A) \leq \nu(A) \leq \sum_i \nu(A \cap A_i) \leq \sum_i \nu(A_i).$$

Taking the infimum over all such covers, we get $\nu(A) = \mu^*(A)$ and μ is indeed an extension of ν. ■

Remark
If the measures are probability measures, then we have that the extension is unique by corollary 1.19.

Theorem 1.25
If F is non-decreasing, right-continuous and satisfies that $F(-\infty) = 0$, $F(\infty) = 1$, then there is a unique probability measure such that

$$\mathrm{P}((-\infty, x]) = F(x).$$

Proof. Define

$$\mathcal{S} = \{(a, b] \mid -\infty \leq a < b \leq \infty\} \cup \{\varnothing\}.$$

It is clear that $\mathcal{S}$ is a semi-algebra. Define the set function $\mathrm{P} : \mathcal{S} \to [0,1]$ by

$$\mathrm{P}((a,b]) = F(b) - F(a)$$

and $\mathrm{P}(\varnothing) = 0$. For disjoint, at most countable $(a_i, b_i] \in \mathcal{S}$, we define

$$\mathrm{P}(\cup_i (a_i, b_i]) = \sum_i \mathrm{P}((a_i, b_i]) = \sum_i F(b_i) - F(a_i).$$

It is clear that P is finitely additive. If $(a,b] = \cup_i (a_i, b_i]$ for disjoint $(a_i, b_i] \in \mathcal{S}$, we may assume without loss of generality that $a = a_1 < b_1 < b_2 < \cdots < b_n = b$ and

$$\mathrm{P}((a,b]) = F(b) - F(a) = \sum_i F(b_i) - F(a_i) = \sum_i \mathrm{P}((a_i, b_i]).$$

Hence P is σ-additive. It now follows from the Carathéodory extension theorem that P can be extended uniquely to a probability measure on $\sigma(\mathcal{S}) = \mathcal{B}$. ■

Remark
This theorem shows that the distribution function completely characterizes the probability measure. In other words, the term "distribution function" can refer to either the CDF or the probability measure.

1.2. Random Variable

Definition 1.26
Let Ω be a probability space. A ***random variable*** *X is a measurable function $X : \Omega \to (S, \mathcal{S})$, where $(S, \mathcal{S})$ is a measurable space.*

Remark
The codomain is often taken to be $(\mathbb{R}, \mathcal{B})$ or $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, but it is also possible to define random functions, i.e., $(S, \mathcal{S})$ is a function space.

Definition 1.27
Let $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ be a random variable. The ***distribution*** *of X is the pushforward measure of P under X, i.e.,*

$$\mu_X(A) = \mathrm{P}(X \in A) = \mathrm{P}(X^{-1}(A)), \quad A \in \mathcal{S}.$$

Definition 1.28
Let $X : (\Omega, \mathcal{F}) \to (\mathbb{R}^d, \mathcal{B})$ be a random variable. The ***cumulative distribution function*** *of X is defined as*

$$F_X(x) = \mathrm{P}(X \le x) = \mathrm{P}(X_1 \le x_1, \ldots, X_d \le x_d), \quad x = (x_1, \ldots, x_d) \in \mathbb{R}^d.$$

Proposition 1.29
Let $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ be a random variable and F be its cumulative distribution function. Then,

(a) F is non-decreasing, i.e., $x \le y$ implies $F(x) \le F(y)$;

(b) $F(-\infty) = 0$ and $F(\infty) = 1$;

(c) F is right-continuous, i.e., $\lim_{y \to x^+} F(y) = F(x)$;

(d) $F(x^-) = \mathrm{P}(X < x)$;

(e) $\mathrm{P}(X = x) = F(x) - F(x^-)$.

Proof. (a) comes from that $\{X \le x\} \subset \{X \le y\}$ for $x \le y$.

Take $a_n \to \infty$. Then $\{X \le a_n\} \nearrow \Omega$ and $\{X \le -a_n\} \searrow \emptyset$. By theorem 1.6, we have that

$$F(a_n) = \mathrm{P}(X \le a_n) \to \mathrm{P}(\Omega) = 1, \quad F(-a_n) = \mathrm{P}(X \le -a_n) \to \mathrm{P}(\emptyset) = 0.$$

(c) is similar to (b). Take $y_n \to x^+$, then $\{X \le y_n\} \searrow \{X \le x\}$. By theorem 1.6, we have that

$$F(y_n) = \mathrm{P}(X \le y_n) \to \mathrm{P}(X \le x) = F(x).$$

For (d), take $x_n \to x^-$, then $\{X \le x_n\} \nearrow \{X < x\}$. By theorem 1.6, we have that

$$F(x_n) = \mathrm{P}(X \le x_n) \to \mathrm{P}(X < x).$$

For (e), $P(X = x) = P(X \le x) - P(X < x) = F(x) - F(x^-)$. ■

Theorem 1.30
Let F be a non-decreasing, right-continuous function satisfying that $F(-\infty) = 0$ and $F(\infty) = 1$. Then there is a random variable X such that

$$F(x) = \mu_X((-\infty, x]).$$

Proof. Put $\Omega = (0,1)$, $\mathcal{F} = \mathcal{B}$, P be the Lebesgue measure and $X(\omega) = \sup\{x \mid F(x) < \omega\}$. Notice that

$$\begin{aligned}\{X \le x\} &= \{\omega \in \Omega \mid \sup\{y \mid F(y) < \omega\} \le x\} \\ &= \{\omega \in \Omega \mid \text{for all } y > x, F(y) \ge \omega\} \\ &= \{\omega \in \Omega \mid F(x) \ge \omega\}.\end{aligned}$$

Hence $P(X \le x) = P(\{\omega \in \Omega \mid \omega \le F(x)\}) = F(x)$. ■

Definition 1.31
If X and Y are random variables mapping to some measurable space $(S, \mathcal{S})$, then X and Y are said to be ***equal in distribution*** *if $\mu_X = \mu_Y$, denoted by $X \overset{d}{=} Y$.*

Definition 1.32
Let $X : \Omega \to \mathbb{R}$ be a random variable with distribution F. $f : \mathbb{R} \to \mathbb{R}$ is said to be the ***density*** *of X if*

$$F(x) = \int_{-\infty}^{x} f(y)dy$$

for all $x \in \mathbb{R}$.

Remark
If f and g are both densities of X, then $f = g$ a.e.

Remark
If $\mu_X \ll \lambda$, where λ is the Lebesgue measure, then by Radon-Nikodym theorem, there is a density f such that

$$\mu_X(A) = \int_A f(x)d\lambda(x)$$

for all $A \in \mathcal{B}$. Or equivalently, F is absolutely continuous.

Example
Not all random variables have densities, even when its CDF is continuous. Consider the

Cantor function

$$F(x) = \begin{cases} \sum_n \frac{a_n}{2^n}, & x = \sum_n \frac{2a_n}{3^n} \in C \text{ for some } \{a_n\} \in \{0,1\}^{\mathbb{N}} \\ \sup_{y \le x, y \in C} F(y), & x \in [0,1] - C \\ 0, & x < 0 \\ 1, & x > 1 \end{cases}$$

where C is the Cantor set. Then F is a valid CDF, but has no density.

Definition 1.33

A probability measure P *is said to be* ***discrete*** *if there is a countable set S such that* $\mathrm{P}(S^c) = 0$. *A random variable X is said to be* ***discrete*** *if its distribution is.*

Theorem 1.34

Suppose $X : (\Omega, \mathcal{F}) \to (S, \sigma(\mathcal{A}))$ and $\mathcal{A}$ is a collection of subsets in S. If $X^{-1}(A) \in \mathcal{F}$ for all $A \in \mathcal{A}$, then X is a random variable.

Proof. Set $\mathcal{G} = \left\{A \subset S \,\middle|\, X^{-1}(A) \in \mathcal{F}\right\}$. Clearly $\varnothing \in \mathcal{G}$ and if $A \in \mathcal{G}$, $X^{-1}(A^c) = (X^{-1}(A))^c \in \mathcal{F}$, so $A^c \in \mathcal{G}$. If $A_n \in \mathcal{G}$, then $X^{-1}(\cup_n A_n) = \cup_n X^{-1}(A_n) \in \mathcal{F}$, so $\cup_n A_n \in \mathcal{G}$. Hence $\mathcal{G}$ is a σ-algebra containing $\mathcal{A}$, so $\sigma(\mathcal{A}) \subset \mathcal{G}$. It follows that $X^{-1}(B) \in \mathcal{F}$ for all $B \in \sigma(\mathcal{A})$, so X is a random variable. ∎

Corollary 1.35

If X_i are random variables, then

$$\inf_i X_i, \quad \sup_i X_i, \quad \liminf_{i\to\infty} X_i, \quad \limsup_{i\to\infty} X_i$$

are all random variables.

Proof. Since the sets of the form $(-\infty, x]$ generate $\mathcal{B}$, it suffices to check that the inverse images of these sets are in $\mathcal{F}$. For $\inf_i X_i$,

$$\left\{\inf_i X_i \le x\right\} = \cup_i \{X_i \le x\} \in \mathcal{F}.$$

For $\sup_i X_i$, since $\sup_i X_i = -\inf_i(-X_i)$, it is also a random variable. Finally, write

$$\liminf_i X_i = \sup_n \inf_{i \ge n} X_i, \quad \limsup_i X_i = \inf_n \sup_{i \ge n} X_i.$$

The results follow from the measurability of $\inf_i X_i$ and $\sup_i X_i$. ∎

Definition 1.36

Let X be a random variable. $\sigma(X)$ is the smallest σ-algebra such that X is measurable.

Remark

If $X : \Omega \to (S, \mathcal{S})$, then $\sigma(X) = X^{-1}(\mathcal{S})$.

Definition 1.37
*Let X be a random variable. The **expectation** of X is defined as*

$$\mathrm{E}[X] = \int X d\,\mathrm{P}.$$

Theorem 1.38 (Jensen's Inequality)
Let $X : \Omega \to \mathbb{R}^d$ be a random variable such that $\mathrm{E}[\|X\|_1] < \infty$ and $\phi : \mathbb{R}^d \to \mathbb{R}$ be a convex function. Then

$$\phi(\mathrm{E}[X]) \le \mathrm{E}[\phi(X)].$$

Proof. For any given $y \in \mathbb{R}^d$, note that $\left\{x \in \mathbb{R}^d \,\middle|\, \phi(x) > \phi(y)\right\}$ is a open convex set. By the Hahn-Banach seperation theorem, there is a hyperplane $\{f(x) = a + \langle b, x\rangle\}$ separating $\left\{(x, \phi(x)) \in \mathbb{R}^{d+1} \,\middle|\, \phi(x) > \phi(y)\right\}$ and $\{(y, \phi(y))\}$. Note that $\phi(y) = f(y)$ and $\phi(x) \ge f(x)$ for all $x \in \mathbb{R}^d$. Take $y = \mathrm{E}[X]$, then

$$\phi(\mathrm{E}[X]) = f(\mathrm{E}[X]) = \mathrm{E}[f(X)] \le \mathrm{E}[\phi(X)].$$

■

Theorem 1.39 (Hölder's Inequality)
Let X, Y be random variables and $p, q \in [1, \infty]$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Then

$$\mathrm{E}[|XY|] \le \mathrm{E}[|X|^p]^{1/p}\,\mathrm{E}[|Y|^q]^{1/q}.$$

Proof. If $\mathrm{E}[|X|^p]$ and $\mathrm{E}[|Y^q|]$ are zero or infinite, the result is trivial. We assume that $\mathrm{E}[|X|^p] = \mathrm{E}[|Y|^q] = 1$. For fixed $y \ge 0$, set $\phi(x) = x^p/p + y^p/p - xy$ for $x \ge 0$.

$$\phi'(x) = x^{p-1} - y, \quad \phi''(x) = (p-1)x^{p-2} \ge 0.$$

Thus ϕ is convex and minimized at $x = y^{1/(p-1)}$ with minimum $\phi(y^{1/(p-1)}) = 0$. Hence $x^p/p + y^q/q \ge xy$ for all $x, y \ge 0$.

$$\mathrm{E}[|XY|] \le \mathrm{E}\left[\frac{|X|^p}{p} + \frac{|Y|^q}{q}\right] = \frac{1}{p} + \frac{1}{q} = 1 = \mathrm{E}[|X|^p]^{1/p}\,\mathrm{E}[|Y|^q]^{1/q}.$$

■

Theorem 1.40 (Markov's Inequality)
If $X \ge 0$ is a random variable, then for any $c > 0$,

$$\mathrm{P}(X \ge c) \le \frac{1}{c}\,\mathrm{E}[X].$$

Proof.

$$\mathrm{P}(X \ge c) = \int \mathbf{1}\{X \ge c\}\, d\,\mathrm{P} \le \int \frac{X}{c} d\,\mathrm{P} = \frac{1}{c}\,\mathrm{E}[X].$$

■

Example
Suppose $\phi : \mathbb{R} \to \mathbb{R}$ is a non-negative function. Put

$$I_A = \inf_{y \in A} \phi(y),$$

where A is some measurable set. Then for any random variable X,

$$I_A \mathbf{1}\{X \in A\} \le \phi(x)\mathbf{1}\{X \in A\} \le \phi(x).$$

Thus

$$I_A \, \mathrm{P}(X \in A) \le \mathrm{E}[\phi(X)].$$

Corollary 1.41 (Chebyshev's Inequality)
Let X be a random variable. Then for any $c > 0$ and $\alpha \in \mathbb{R}$,

$$\mathrm{P}(|X - \alpha| \ge c) \le \frac{1}{c^2} \mathrm{E}\left[(X - \alpha)^2\right].$$

Proof. By the Markov's inequality,

$$\mathrm{P}(|X - \alpha| \ge c) = \mathrm{P}((X - \alpha)^2 \ge c^2) \le \frac{1}{c^2} \mathrm{E}\left[(X - \alpha)^2\right].$$

■

Theorem 1.42
Suppose X is a random variable of $(S, \mathcal{S})$ with distribution μ and $f : (S, \mathcal{S}) \to (\mathbb{R}, \mathcal{B})$ is measurable. If either

- *(a)* $f \ge 0$, *or*
- *(b)* $\mathrm{E}[|f(X)|] < \infty$,

then

$$\mathrm{E}[f(X)] = \int f(x) d\mu(x).$$

Proof. Suppose first that $f = \mathbf{1}_A$ for some $A \in \mathcal{S}$. Then

$$\mathrm{E}[f(X)] = \mathrm{P}(X \in \mathcal{A}) = \mathrm{P}(X^{-1}(A)) = \mu(A) = \int \mathbf{1}_A d\mu.$$

By linearity we can extend this result to simple functions. Now suppose first that (a) holds. For such f, there is a sequence of simple functions $s_n \nearrow f$ and $s_n \circ X \nearrow f \circ X$. By LMCT,

$$\mathrm{E}[f(X)] = \mathrm{E}\left[\lim_n s_n(X)\right] = \lim_n \mathrm{E}[s_n(X)] = \lim_n \int s_n d\mu = \int f d\mu.$$

Suppose that (b) is the case. Write $f = f^+ - f^-$ and apply the previous result.

$$\mathrm{E}[f(X)] = \mathrm{E}\left[f^+(X)\right] - \mathrm{E}[f^-(X)] = \int f^+ d\mu - \int f^- d\mu = \int f d\mu.$$

■

Definition 1.43
*The k-**th moment** of a random variable X is* $\mathrm{E}\left[X^k\right]$.

Definition 1.44
*The **variance** of a random variable X is* $\mathrm{Var}\,\mathrm{E}\left[(X - \mathrm{E}[X])^2\right]$.

Definition 1.45
*The **covariance** of two integrable random variables X, Y is*

$$\mathrm{Cov}(X,Y) = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])].$$

1.3. Independence

Definition 1.46

*Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space. Suppose $\mathcal{F}_\beta \subset \mathcal{F}$, $\beta \in B$ are a collection of sub-σ-algebras. Then $\{\mathcal{F}_\beta\}$ are **independent** if for all finite $\{\mathcal{F}_i\}_{i=1}^n \subset \{\mathcal{F}_\beta\}$,*

$$\mathrm{P}(\cap_{i=1}^n A_i) = \prod_{i=1}^n \mathrm{P}(A_i)$$

where $A_i \in \mathcal{F}_i$.

Definition 1.47

*A collection of random variables $\{X_\beta \mid \beta \in B\}$ on $(\Omega, \mathcal{F}, \mathrm{P})$ is **independent** if the collection of the generating σ-algebras $\{\sigma(X_\beta) \mid \beta \in B\}$ is.*

Remark

In other words,

$$\mathrm{P}(\cap_i \{X_{\beta_i} \in A_i\}) = \prod_i \mathrm{P}(X_{\beta_i} \in A).$$

Note that these random variables can map into different measurable space.

Definition 1.48

*A collection of events $\mathcal{S}$ is **independent** if $\{\mathbf{1}_A \mid A \in \mathcal{S}\}$ is.*

Proposition 1.49

Let $X_1, \ldots, X_n$ be independent random variables and $g_1, \ldots g_n$ are measurable functions. Then $g_1(X_1), \ldots, g_n(X_n)$ are independent.

Proof. Suppose $g_i : (S_i, \mathcal{S}_i) \to (T_i, \mathcal{T}_i)$. For $A_i \in \mathcal{T}_i$, $g^{-1}(A_i) \in \mathcal{S}_i$ and

$$\mathrm{P}(\cap_i \{g_i(X_i) \in A_i\}) = \mathrm{P}(\cap_i \{X_i \in g^{-1}(A_i)\}) = \prod_i \mathrm{P}(X_i \in g^{-1}(A_i)) = \prod_i \mathrm{P}(g_i(X_i) \in A_i).$$

$g_1(X_1), \ldots, g_n(X_n)$ are independent. ■

Theorem 1.50

Let $\mathcal{S}_1, \ldots \mathcal{S}_n$ be a collection of π-system. If $\Omega \in \mathcal{S}_i$ for all $i = 1, \ldots, n$ and for all $A_i \in \mathcal{S}_i$,

$$\mathrm{P}(\cap_i A_i) = \prod_i \mathrm{P}(A_i),$$

then $\sigma(\mathcal{S}_1), \ldots, \sigma(\mathcal{S}_n)$ are independent.

Proof. Fix $\mathcal{S}_2, \ldots, \mathcal{S}_n$. Put

$$\mathcal{L} = \left\{ A \in \mathcal{F} \;\middle|\; \mathrm{P}(A \cap (\cap_{i=2}^n A_i)) = \mathrm{P}(A) \prod_{i=2}^n \mathrm{P}(A_i), A_i \in \mathcal{S}_i \text{ for } i = 2, \ldots, n \right\}.$$

We claim that $\mathcal{L}$ forms a λ-system. First, by assumption we can pick $A_i = \Omega$ for $i = 2, \dots, n$ to see that $\Omega \in \mathcal{L}$. Suppose that $A \subset B$, $A, B \in \mathcal{L}$,

$$\begin{aligned} P((B - A) \cap (\cap_{i=2}^n A_i)) &= P((B \cap (\cap_{i=2}^n A_i)) - (A \cap (\cap_{i=2}^n A_i))) \\ &= P(B) \prod_{i=2}^n P(A_i) - P(A) \prod_{i=2}^n P(A_i) = P(B - A) \prod_{i=2}^n P(A_i). \end{aligned}$$

Hence $B - A \in \mathcal{L}$. Let $S_j \nearrow S$, $S_j \in \mathcal{L}$. Then

$$P(S \cap (\cap_{i=2}^n A_i)) = \lim_{j\to\infty} P(S_j \cap (\cap_{i=2}^n A_i)) = \lim_{j\to\infty} P(S_j) \prod_{i=2}^n P(A_i) = P(S) \prod_{i=2}^n P(A_i).$$

Thus $S \in \mathcal{L}$ and $\mathcal{L}$ is a λ-system. By Dynkin's π-λ, $\sigma(\mathcal{S}_1), \mathcal{S}_2, \dots, \mathcal{S}_n$ satisfies the product property. Repeat the procedure for $\mathcal{S}_2, \dots, \mathcal{S}_n$. We have that $\sigma(\mathcal{S}_1), \dots, \sigma(\mathcal{S}_n)$ satisfies the product property. That is, they are independent. ■

Corollary 1.51
Let $X_1, \dots, X_n$ be $\mathbb{R}$-valued random variables. Then they are independent if and only if

$$P(X_1 \le s_1, \dots, X_n \le s_n) = \prod_{i=1}^n P(X_i \le s_i)$$

for all $s_i \in \mathbb{R}$, $1 \le i \le n$.

Proof. The sufficient part is trivial. For the converse, put $\mathcal{S}_i = \{\{X_i \le t\} \mid t \in \mathbb{R}\} \cup \{\Omega\}$. Clearly $\mathcal{S}_i$ are π-system and $\Omega \in \mathcal{S}_i$ for all i. $\sigma(\mathcal{S}_i)$ are independent and $\mathcal{S}_i$ generates $\sigma(X_i)$. Applying theorem 1.50 shows that X_i are independent. ■

Corollary 1.52
If $\mathcal{F}_{ij}$, $1 \le i \le n, 1 \le j \le m(i)$ are independent σ-algebras, then $\mathcal{G}_i = \sigma(\cup_j \mathcal{F}_{ij})$ are independent.

Proof. Put $\mathcal{H}_i = \left\{\cap_j A_j \,\middle|\, A_j \in \mathcal{F}_{ij}\right\}$. We claim that $\sigma(\mathcal{H}_i) = \mathcal{G}_i$. Indeed, by choosing sets of the form

$$(\Omega, \dots, \Omega, A_j, \Omega, \dots, \Omega) \in \mathcal{F}_{i1} \times \dots \times \mathcal{F}_{im(i)},$$

it is clear that $\cup_j \mathcal{F}_{ij} \subset \mathcal{H}_i$. Also, if $A \in \mathcal{H}_i$, then

$$A = \cap_j A_j = (\cup_j (A_j^c))^c \in \sigma(\cup_j \mathcal{F}_{ij}).$$

Thus $\cup_j \mathcal{F}_{ij} \subset \mathcal{H}_i \subset \sigma(\cup_j \mathcal{F}_{ij})$ and $\sigma(\mathcal{H}_i) = \sigma(\cup_j \mathcal{F}_{ij}) = \mathcal{G}_i$. Also notice that $\mathcal{H}_i$ contain Ω and form π-systems. For $A_i \in \mathcal{H}_i$, write $A_i = \cap_j A_{ij}$. Then

$$P(\cap_i A_i) = P(\cap_{ij} A_{ij}) = \prod_{ij} P(A_{ij}) = \prod_i P(\cap_j A_{ij}) = \prod_i P(A_i).$$

From theorem 1.50 we know that $\mathcal{G}_i = \sigma(\mathcal{H}_i)$ are independent. ■

Corollary 1.53
If $X_{ij}, 1 \le i \le n, 1 \le j \le m(i)$ are independent random variables, then $Y_i = h_i(X_{i1}, \dots, X_{im(i)})$ are independent provided that h_i are measurable.

Proof. Write $\mathcal{F}_{ij} = \sigma(X_{ij})$. We claim that $\sigma(Y_i) \subset \sigma(\cup_j \mathcal{F}_{ij})$. Indeed, if B_i is a measurable set, $h_i^{-1}(B_i)$ is measurable. Write $h_i^{-1}(B_i) = C_{i1} \times \cdots \times C_{im(i)}$ and since each $X_{ij}^{-1}(C_{ij}) \in \mathcal{F}_{ij}$, we see that $\sigma(Y_i) \subset \sigma(\cup_j \mathcal{F}_{ij})$. It then follows from corollary 1.52 that $\sigma(Y_i)$ are independent and Y_i are independent. ■

Theorem 1.54
If $X_1, \dots X_n$ are independent $\mathbb{R}$-valued random variables and the distribution of X_i is μ_i. Then the joint distribution of $(X_1, \dots, X_n)$ is $\mu_1 \times \cdots \times \mu_n$.

Proof. Let μ be the distribution of $(X_1, \dots, X_n)$. By definition,

$$\begin{aligned}\mu((X_1, \dots) \in A_1 \times \cdots \times A_n) &= \mu(X_1 \in A_1, \dots, X_n \in A_n) \\ &= \prod_{i=1}^{n} \mu_i(X_i \in A_i) = (\mu_1 \times \cdots \times \mu_n)(A_1 \times \cdots \times A_n).\end{aligned}$$

Now the sets of the forms $A = A_1 \times \cdots \times A_n$ is a π-system generating the product σ-algebra. By corollary 1.19, the joint distribution is exactly $\mu_1 \times \cdots \times \mu_n$. ■

Theorem 1.55
Let X, Y be two independent random variables. If $h(x, y)$ satisfies either

(a) $\mathrm{E}[|h(X, Y)|] < \infty$, *or*

(b) *h is non-negative,*

then

$$\mathrm{E}[h(X, Y)] = \int \int h d\mu_X d\mu_Y,$$

where μ_X, μ_Y are the distributions of X and Y, respectively.

Proof. The proof follows directly from Fubini-Tonelli theorem. If one of the assumptions is true, then

$$\mathrm{E}[h(X, Y)] = \int_{\mathbb{R}^2} h d(\mu_X \times \mu_Y) = \int \int h d\mu_X d\mu_Y.$$

■

Remark
If $h(x, y) = h_1(x) h_2(y)$, then

$$\mathrm{E}[h_1(X) h_2(Y)] = \mathrm{E}[h(X, Y)] = \int \int h_1 h_2 d\mu_X d\mu_Y = \mathrm{E}[h_1(X)] \, \mathrm{E}[h_2(Y)].$$

Corollary 1.56
If $X_1, \ldots X_n$ are independent random variables and

(a) $\mathrm{E}\left[|X_1 \cdots X_n|\right] < \infty$ *or*

(b) $X_i \geq 0$ *for all i,*

then

$$\mathrm{E}\left[X_1 \cdots X_n\right] = \prod_{i=1}^{n} \mathrm{E}\left[X_i\right].$$

Proof. Let $h(x,y) = xy$. By assumptions, we have either $\mathrm{E}\left[|h(X_1,X_2)|\right] < \infty$ or $h(X_1,X_2) \geq 0$. By theorem 1.55, $\mathrm{E}\left[X_1X_2\right] = \mathrm{E}\left[X_1\right]\mathrm{E}\left[X_2\right]$. Substitute X_1 by X_1X_2 and X_2 by X_3, we see that $\mathrm{E}\left[X_1X_2X_3\right] = \mathrm{E}\left[X_1\right]\mathrm{E}\left[X_2\right]\mathrm{E}\left[X_3\right]$. Repeat the procedure n times and the result follows. ∎

Definition 1.57
Let X, Y be independent random variables with CDF F and G, respectively. The ***convolution*** *of two CDF is defined as*

$$(F * G)(z) = \int F(z-y)dG(y).$$

Remark
If F and G are absolutely continuous with respect to the Lebesgue measure, then they have Radon-Nikodym derivatives f and g. The definition of convolution becomes

$$(F * G)(z) = \int F(z-y)dG(y) = \int_{-\infty}^{\infty}\int_{-\infty}^{z-y} f(x)g(y)dxdy.$$

Then

$$(F * G)'(z) = \int f(z-y)g(y)dy = (f * g)(z),$$

which is exactly the definition of convolution of two functions.

Proposition 1.58
Let X and Y be independent random variables. Then

$$\mathrm{P}(X + Y \leq z) = (F * G)(z).$$

Proof. By theorem 1.55,

$$\begin{aligned}\mathrm{P}(X + Y \leq z) = \mathrm{E}\left[\mathbf{1}\left\{X + Y \leq z\right\}\right] &= \int\int \mathbf{1}\left\{x + y \leq z\right\} dF(x)dG(y) \\ &= \int F(z-y)dG(y) = (F * G)(z).\end{aligned}$$

∎

Remark
Note that the convolution is commutative since

$$(F * G)(z) = \mathrm{P}(X + Y \leq z) = \mathrm{P}(Y + X \leq z) = (G * F)(z).$$

Remark

For discrete X and Y, the convolution becomes

$$\mathrm{P}(X+Y=z)=\sum_{y}\mathrm{P}(X=z-y)\,\mathrm{P}(Y=y).$$

Example

Consider $X\sim\Gamma(\alpha_1,\beta)$ and $Y\sim\Gamma(\alpha_2,\beta)$. Then the density for $X+Y$ is

$$\begin{aligned}
f_{X+Y}(z) &= \int f_X(z-y)f_Y(y)dy \\
&= \int_0^z \frac{1}{\Gamma(\alpha_1)}\beta^{\alpha_1}(z-y)^{\alpha_1-1}e^{-\beta(z-y)}\frac{1}{\Gamma(\alpha_2)}\beta^{\alpha_2}y^{\alpha_2-1}e^{-\beta y}dy \\
&= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\beta^{\alpha_1+\alpha_2}e^{-\beta z}\int_0^z (z-y)^{\alpha_1-1}y^{\alpha_2-1}dy \\
&= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\beta^{\alpha_1+\alpha_2}e^{-\beta z}z^{\alpha_1+\alpha_2-1}\int_0^1 (1-t)^{\alpha_1-1}t^{\alpha_2-1}dt \\
&= \frac{B(\alpha_1,\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\beta^{\alpha_1+\alpha_2}e^{-\beta z}z^{\alpha_1+\alpha_2-1} = \frac{1}{\Gamma(\alpha_1+\alpha_2)}\beta^{\alpha_1+\alpha_2}e^{-\beta z}z^{\alpha_1+\alpha_2-1}.
\end{aligned}$$

Hence $X+Y\sim\Gamma(\alpha_1+\alpha_2,\beta)$.

1.4. Convergence of Random Variables

Definition 1.59
A sequence of probability measures P_n *on* $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ *are* ***consistent*** *if*

$$\mathrm{P}_{n+1}((a_1, b_1] \times \cdots \times (a_n, b_n] \times \mathbb{R} = \mathrm{P}_n((a_1, b_1] \times \cdots \times (a_n, b_n])$$

for every n*.*

Theorem 1.60 (Kolmogorov Extension)
Suppose that a sequence of probability measures P_n *on* $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ *are consistent. Then there is a unique probability measure* P *on* $\mathbb{R}^{\mathbb{N}}$ *satisfying that*

$$\mathrm{P}(\{\omega \in \mathbb{R}^{\mathbb{N}} \mid \omega_i \in (a_i, b_i], 1 \le i \le n\}) = \mathrm{P}_n((a_1, b_1] \times \cdots \times (a_n, b_n]).$$

Proof. Let

$$\mathcal{S} = \{(a_1, b_1] \times \cdots \times (a_n, b_n] \times \mathbb{R} \times \cdots \mid n \in \mathbb{N}\}.$$

Define P on $\mathcal{S}$ to be

$$\mathrm{P}((a_1, b_1] \times \cdots \times (a_n, b_n] \times \mathbb{R} \times \cdots) = \mathrm{P}_n((a_1, b_1] \times \cdots \times (a_n, b_n])$$

Clearly, $\mathcal{S}$ forms a semi-algebra. From the Carathéodory extension theorem, it suffices to show that P is finitely additive, σ-additive on $\mathcal{S}$ and $\mathrm{P}(\varnothing) = 0$. Note that $\mathrm{P}(\varnothing) = \mathrm{P}(\varnothing \times \mathbb{R} \times \cdots) = \mathrm{P}_1(\varnothing) = 0$. We verify the first two conditions.

First, if $A, B \in \mathcal{S}$ are disjoint, $m \le n$,

$$A = \{\omega \in \mathbb{R}^{\mathbb{N}} \mid \omega_i \in (a_i, b_i], 1 \le i \le m\} \quad \text{and} \quad B = \{\omega \in \mathbb{R}^{\mathbb{N}} \mid \omega_i \in (c_i, d_i], 1 \le i \le n\},$$

then

$$\mathrm{P}(A \cup B) = \mathrm{P}_n((\pi_n A) \cup (\pi_n B)) = \mathrm{P}_n(\pi_n A) + \mathrm{P}_n(\pi_n B) = \mathrm{P}(A) + \mathrm{P}(B),$$

where $\pi_n : \omega \to (\omega_1, \ldots, \omega_n)$ is the projection onto the first n components. Hence P is finitely additive.

Next, suppose $A_1, \ldots \in \mathbb{R}^{\mathbb{N}}$ are countably many disjoint measurable sets. Put $A = \cup_i A_i$. We can consider the algebra $\bar{\mathcal{S}} = \{\text{finite disjoint union of sets in } \mathcal{S}\}$ generated by $\mathcal{S}$. $B_n = \cup_{i=1}^n A_i \in \bar{\mathcal{S}}$. Thus

$$\mathrm{P}(A) = \mathrm{P}(B_n) + \sum_{i=1}^{n} \mathrm{P}(A_n)$$

by the previous result. It now suffices to show that $\mathrm{P}(B_n) \to 0$ for any $B_n \searrow \varnothing$. ∎