# I. Probability Measures

## 1. $\sigma$-fields

Def'n 1.1. $\sigma$-**field** of Subsets of $\Omega$
Let $\Omega$ be a set and let $\mathcal{F} \subseteq 2^{\Omega}$. We say $\mathcal{F}$ is a ***$\sigma$-field*** of subsets of $\Omega$ if

(a) $\Omega \in \mathcal{F}$;
(b) $A \in \mathcal{F}$ implies $\Omega \setminus A \in \mathcal{F}$; and *closure under complement*
(c) $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$ implies $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$. *closure under countable union*

The elements of $\mathcal{F}$ are called ***events*** and the pair $(\Omega, \mathcal{F})$ is called a measurable space.

**Example 1.1.** Simple $\sigma$-fields

Let $\Omega$ be a set.

(a) The ***trivial $\sigma$-field*** is $\{\emptyset, \Omega\}$.
(b) The power set $2^{\Omega}$ is also a $\sigma$-field.
(c) Given any $A \subseteq \Omega$, $\{\emptyset, A, \Omega \setminus A, \Omega\}$ is a $\sigma$-field.
(d) The collection of countable and co-countable sets,

$$\mathcal{F} = \{A \subseteq \Omega : A \text{ is countable or } X \setminus A \text{ is countable}\},$$

is a $\sigma$ field. To see this, let $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$. If every $A_n$ is countable, then so is $\bigcup_{n=1}^{\infty} A_n$. Hence $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.
On the other hand, if any $A_m$ is co-countable, then

$$X \setminus \left(\bigcup_{n=1}^{\infty} A_n\right) = X \setminus \left(A_m \cup \bigcup_{n=1, n\neq m}^{\infty} A_n\right) = \left(X \setminus \left(\bigcup_{n=1, n\neq m}^{\infty} A_n\right)\right) \setminus A_m \subseteq X \setminus A_m,$$

so that $X \setminus \left(\bigcup_{n=1}^{\infty} A_n\right)$ is co-countable. Thus $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

---

**Example 1.2.** An Example of Important Non-$\sigma$-field

Consider $\Omega = [0,1]$ and consider

$$\begin{aligned}\mathcal{B}_0 &= \{\text{the finite unions of disjoint left-open-right-closed intervals}\} \\ &= \left\{\bigcup_{i=1}^{n} (a_i, b_i] : (a_i, b_i] \subseteq [0,1], (a_i, b_i] \cap (a_j, b_j] = \emptyset \text{ for } i \neq j\right\}.\end{aligned}$$

Then $\mathcal{B}_0$ is not a $\sigma$-field, since given $A_1 = \left(\frac{1}{2}, 1\right], A_2 = \left(\frac{1}{3}, \frac{1}{2}\right], \ldots, A_n = \left(\frac{1}{n+1}, \frac{1}{n}\right], \ldots \subseteq [0,1]$, we obtain:

$$\bigcup_{n=1}^{\infty} A_n = [0,1] \notin \mathcal{B}_0\,.$$

However, we can check that the first two axioms ($\Omega \in \mathcal{B}_0$ and closure under complement) hold for $\mathcal{B}_0$ and that $\mathcal{F}$ is closed under *finite*, but not countable, union.

$\mathcal{B}_0$ is an example of ***field*** (or ***algebra***) of subsets of $\Omega$.

---

Def'n 1.2. $\sigma$-field **Generated** by a Collection
Let $\Omega$ be a set and let $\mathcal{A} \subseteq 2^{\Omega}$. If we let

$$\sigma(\mathcal{A}) = \bigcap_{\substack{\mathcal{F} \supseteq \mathcal{A}: \\ \mathcal{F} \text{ is a } \sigma\text{-field}}} \mathcal{F},$$

then $\sigma(\mathcal{A})$ is a $\sigma$-field, called the $\sigma$-field ***generated*** by $\mathcal{A}$.

**Example 1.3.** Generating $\sigma$-fields

Let $\Omega$ be a set.

(a) The trivial $\sigma$-field is generated by $\emptyset$.

(b) $\{\emptyset, A, \Omega \setminus A, \Omega\}$ is generated by $\{\Omega \setminus A\}$.

(c) The $\sigma$-field of countable and co-countable sets $\mathcal{F}$ is generated by $\{\{\omega\}\}_{\omega\in\Omega}$, the collection of singletons.

---

**Example 1.4.** Borel $\sigma$-field on $(0, 1]$

The ***Borel $\sigma$-field*** on $(0, 1]$ is the $\sigma$-field generated by $\mathcal{B}_0$ (see Example 1.2).

It can also be generated by $\{[a, b] \subseteq \Omega\}$, $\{(a, b) \subseteq \Omega\}$, $\{(a, b] \subseteq \Omega\}$, $\{[a, b) \subseteq \Omega\}$ (exercise).

---

**Example 1.5.** General Borel $\sigma$-fields on Topological Spaces

Given any topological space $(\Omega, \tau)$, then the Borel $\sigma$-field on $\Omega$ is defined as $\sigma(\tau)$. If we let $\gamma$ to be the collection of closed sets then $\gamma$ also generates the Borel $\sigma$-fields on $\Omega$.

---

## 2. Probability Measure

Def'n 1.3. **Probability Measure** on a Measurable Space

Let $(\Omega, \mathcal{F})$ be a measurable space. We say $\mathbb{P} : \mathcal{F} \to [0, \infty]$ is a ***measure*** on $\mathcal{F}$ if

(a) $\mathbb{P}(\emptyset) = 0$; and

(b) $\mathbb{P}\left(\dot{\bigcup}_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$. *countable additivity*

If $\mathbb{P}$ satisfy in addition that $\mathbb{P}(A) \in [0, 1]$ for all $A \in \mathcal{F}$, then we say $\mathbb{P}$ is a ***probability measure***.

Def'n 1.4. **Probability Space**

A ***probability space*** is a triplet $(\Omega, \mathcal{F}, \mathbb{P})$ such that

(a) $\Omega$ is a set called ***sample space***, the set of all the possible results of a random experiments or observations;

(b) $\mathcal{F}$ is a ***$\sigma$-field*** of subsets of $\Omega$; and

(c) $\mathbb{P}$ is a ***probability measure*** on $\mathcal{F}$.

**Example 1.6.** Tossing a Coin

When we are tossing a coin $n$ times,

$$\begin{aligned} \Omega &= \{0, 1\}^n, \\ \mathcal{F} &= 2^{\Omega}, \\ \mathbb{P}(A) &= \frac{|A|}{2^n}, \qquad \forall A \in \mathcal{F}. \end{aligned}$$

---

**Example 1.7.** Discrete Probability Space

Let $\Omega$ be a countable set and let $p : \Omega \to [0, 1]$ be such that

$$\sum_{\omega\in\Omega} p(\omega) = 1.$$

Then $\mathbb{P} : 2^{\Omega} \to [0, 1]$ by

$$\mathbb{P}(A) = \sum_{\omega\in A} p(\omega), \qquad \forall A \subseteq \Omega,$$

is a probability measure on $\left(\Omega, 2^{\Omega}\right)$.

We call $\left(\Omega, 2^{\Omega}, \mathbb{P}\right)$ a ***discrete*** probability space.

---

**Proposition 1.1.** Properties of Probability Measures

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Then

(a) $A \subseteq B$ implies $\mathbb{P}(A) \leq \mathbb{P}(B)$; *monotonicity*

(b) $A \subseteq B$ implies $\mathbb{P}(A) = \mathbb{P}(B \setminus A) + \mathbb{P}(B)$; *excision*

(c) given any $\{A_i\}_{i=1}^n \subseteq \mathcal{F}$,

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}\left(A_i \cap A_j\right) + \cdots + (-1)^{n-1}\mathbb{P}(A_1 \cap \cdots \cap A_n);$$

*inclusion-exclusion*

(d) for any increasing chain $(A_n)_{n=1}^\infty \in \mathcal{F}^{\mathbb{N}}$, we have

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_{n=1}^\infty A_n\right);$$

*continuity from below*

(e) for any decreasing chain $(A_n)_{n=1}^\infty \in \mathcal{F}^{\mathbb{N}}$, we have

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcap_{n=1}^\infty A_n\right);$$

*continuity from above*

and

(f) for any $\{A_n\}_{n=1}^\infty \subseteq \mathcal{F}$,

$$\mathbb{P}\left(\bigcup_{n=1}^\infty A_n\right) \leq \sum_{n=1}^\infty \mathbb{P}(A_n).$$

*countable subadditivity (Bool's inequality)*

**Proof.**

(a) Suppose $A \subseteq B$. Then $B \setminus A \in \mathcal{F}$ as well with $B = A \mathbin{\dot\cup} B \setminus A$, so that $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A)$.

(b) This is shown in (a).

(c) When $n = 2$, we have $A_1 \cup A_2 = (A_1 \setminus (A_1 \cap A_2)) \mathbin{\dot\cup} A_2$, so that

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1 \setminus (A_1 \cap A_2)) + \mathbb{P}(A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2).$$

Assume that (c) holds for some $n \geq 2$. Then we note that

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^{n+1} A_i\right) &= \mathbb{P}\left(\left(\bigcup_{i=1}^{n} A_i\right) \cup A_{n+1}\right) \\
&= \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) + \mathbb{P}(A_{n+1}) - \mathbb{P}\left(\left(\bigcup_{i=1}^{n} A_i\right) \cap A_{n+1}\right) \\
&= \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) + \mathbb{P}(A_{n+1}) - \mathbb{P}\left(\bigcup_{i=1}^{n} (A_i \cap A_{n+1})\right) \\
&= \left(\sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}\left(A_i \cap A_j\right) + \cdots + (-1)^{n-1}\mathbb{P}(A_1 \cap \cdots \cap A_n)\right) \\
&\quad + \mathbb{P}(A_{n+1}) - \left(\sum_i \mathbb{P}(A_i \cap A_{n+1}) - \sum_{i<j} \mathbb{P}\left(A_i \cap A_j \cap A_{n+1}\right) + \cdots + (-1)^{n+1}\mathbb{P}(A_1 \cap \cdots A_{n+1})\right) \\
&= \sum_{i=1}^{n+1} \mathbb{P}(A_i) - \sum_{i<j}^{n+1} \mathbb{P}\left(A_i \cap A_j\right) + \cdots + (-1)^{n+1}\mathbb{P}(A_1 \cap \cdots A_{n+1}).
\end{aligned}$$

(d) Let $(A_n)_{n=1}^{\infty} \in \mathcal{F}^{\mathbb{N}}$ be an increasing chain. Define

$$B_n = A_n \setminus \bigcup_{i=1}^{n-1} A_i = A_n \setminus A_{n-1}$$

for all $n \in \mathbb{N}$. Then we observe that each $B_n \in \mathcal{F}$ with $A_n = \dot\bigcup_{i=1}^{n} B_i$, so that $\bigcup_{n=1}^{\infty} A_n = \dot\bigcup_{i=1}^{\infty} B_i$ and

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} \mathbb{P}\left(\dot\bigcup_{i=1}^{n} B_i\right) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(B_i) = \sum_{i=1}^{\infty} \mathbb{P}(B_i) = \mathbb{P}\left(\dot\bigcup_{i=1}^{\infty} B_i\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right),$$

as required.

(e) It suffices to note that, by taking $B_n = \Omega \setminus A_n$ for all $n \in \mathbb{N}$, we have

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} 1 - \mathbb{P}(B_n) = 1 - \mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) = \mathbb{P}\left(\Omega \setminus \bigcup_{n=1}^{\infty} B_n\right) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right).$$

(f) Let

$$B_n = A_n \setminus \bigcup_{i=1}^{n-1} A_i \subseteq A_n$$

for all $n \in \mathbb{N}$. Then $\bigcup_{n=1}^{\infty} A_n = \dot\bigcup_{n=1}^{\infty} B_n$, so that

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathbb{P}\left(\dot\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(B_n) \le \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

**QED**

## 3. Construction of Probability Measures

Given a $\sigma$-field $\mathcal{F}$, it is hard to describe every elements in $\mathcal{F}$. This means it is also hard to assign a probability measure $\mathbb{P}$ on $\mathcal{F}$.

A natural idea to get around this is to first define $\mathbb{P}$ on a subset of $\mathcal{F}$and then extend it to the whole $\mathcal{F}$.

Def'n 1.5. **Field** of Subsets of $\Omega$

We say $\mathcal{F}_0 \subseteq 2^{\Omega}$ is a ***field*** of subsets of $\Omega$ if

(a) $\emptyset \in \mathcal{F}_0$;

(b) $A \in \mathcal{F}_0$ implies $\Omega \setminus A \in \mathcal{F}$; and *closure under complement*

(c) $A, B \in \mathcal{F}_0$ implies $A \cup B \in \mathcal{F}_0$. *closure under finite union*

That is, a field is a subcollection of $2^{\Omega}$ that *looks like* a $\sigma$-field that has closure under *finite* union instead of countable union.

Def'n 1.6. **Premeasure** on a Field

Let $\mathcal{F}_0$ be a field of subsets of $\Omega$. We say $\mathbb{P}_0 : \mathcal{F}_0 \to [0, \infty]$ is a ***premeasure*** on $\mathcal{F}$ if

(a) $\mathbb{P}_0(\emptyset) = 0$; and

(b) for any subcollection $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}_0$ of disjoint elements,

$$\mathbb{P}_0\left(\dot\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}_0(A_n).$$ *countable additivity*

If

(c) $\mathbb{P}_0(\Omega) = 1$

as well, then we say $\mathbb{P}_0$ is a ***probability*** premeasure.

Def'n 1.7. **Outer Measure** on a Set

We say $\mathbb{P}^* : 2^\Omega \to [0, \infty]$ is an ***outer measure*** on $\Omega$ if

(a) $\mathbb{P}^*(\emptyset) = 0$; and

(b) for any $\{A_n\}_{n=1}^\infty \subseteq 2^\Omega$, $\mathbb{P}^*\left(\bigcup_{n=1}^\infty A_n\right) \le \sum_{n=1}^\infty \mathbb{P}^*(A_n)$. *countable subadditivity*

Given $A \subseteq \Omega$, we say $A$ is $\mathbb{P}^*$-***measurable*** if

$$\mathbb{P}^*(E) = \mathbb{P}^*(E \cap A) + \mathbb{P}(E \cap (\Omega \setminus A)), \qquad \forall E \subseteq \Omega. \qquad \textit{Caratheodory's criterion}$$

**Theorem 1.2.** Extension Theorem

Let $\mathcal{F}_0$ be a field of subsets of $\Omega$ and let $\mathbb{P}_0 : \mathcal{F}_0 \to [0,1]$ be a probability premeasure on $\mathcal{F}_0$. Then there exists a unique probability measure $\mathbb{P} : \sigma(\mathcal{F}_0) \to [0,1]$ such that $\mathbb{P}|_{\mathcal{F}_0} = \mathbb{P}_0$.

---

We split the proof of the theorem into few results.

**Proof of Existence.** Let $\mathbb{P}^* : 2^\Omega \to [0,1]$ be defined by

$$\mathbb{P}^*(A) = \inf\left\{\sum_{n=1}^\infty \mathbb{P}(A_n) : \{A_n\}_{n=1}^\infty \subseteq \mathcal{F}_0, A \subseteq \bigcup_{n=1}^\infty A_n\right\}, \qquad \forall A \subseteq \Omega,$$

which is an outer measure on $\Omega$. Then by taking

$$\mathcal{F} = \{A \subseteq \Omega : A \text{ is } \mathbb{P}^*\text{-measurable}\},$$

we know that $\mathcal{F}$ is a $\sigma$-field and $\mathbb{P} = \mathbb{P}^*|_{\mathcal{F}}$ is a probability measure on $(\Omega, \mathcal{F})$ by Caratheodory's theorem.

Now we check few claims.

- *Claim 1.* $\mathcal{F}_0 \subseteq \mathcal{F}$.

  Proof. Let $A \in \mathcal{F}_0$. For any $E \subseteq \Omega$, we desire to show

  $$\mathbb{P}^*(E) = \mathbb{P}^*(E \cap A) + \mathbb{P}^*(E \cap (\Omega \setminus A)).$$

  For any $\varepsilon > 0$, let $\{A_n\}_{n=1}^\infty \subseteq \mathcal{F}_0$ be such that, and

  $$B_n = A \cap A_n, C_n = A \cap (\Omega \setminus A_n), \qquad \forall n \in \mathbb{N}.$$

  Then

  $$\begin{aligned} E \cap A &\subseteq \bigcup_{n=1}^\infty B_n \\ E \cap (\Omega \setminus A) &\subseteq \bigcup_{n=1}^\infty C_n \end{aligned}.$$

  This means

  $$\begin{aligned} \mathbb{P}^*(E \cap A) &\le \mathbb{P}^*\left(\bigcup_{n=1}^\infty B_n\right) \le \sum_{n=1}^\infty \mathbb{P}(B_n) \\ \mathbb{P}^*(E \cap (\Omega \setminus A)) &\le \mathbb{P}^*\left(\bigcup_{n=1}^\infty C_n\right) \le \sum_{n=1}^\infty \mathbb{P}(C_n) \end{aligned}.$$

  Thus

  $$\mathbb{P}^*(E \cap A) + \mathbb{P}^*(E \cap (\Omega \setminus A)) \le \sum_{n=1}^\infty \mathbb{P}(B_n) + \sum_{n=1}^\infty \mathbb{P}(C_n) = \sum_{n=1}^\infty \mathbb{P}(A_n) \le \mathbb{P}^*(E) + \varepsilon.$$

Since this holds for all $\varepsilon > 0$, it follows $\mathbb{P}^*(E) \geq \mathbb{P}^*(E \cap A) + \mathbb{P}^*(E \cap (\Omega \setminus A))$.

The other direction is trivial, as union of any covers for $E \cap A, E \cap (\Omega \setminus A)$, respectively, is a cover for $E$. Hence we have shown the desired equality, which imply $A \in \mathcal{F}$. Thus $\mathcal{F}_0 \subseteq \mathcal{F}$.

A consequence of Claim 1 is that, $\mathcal{F} \supseteq \sigma(\mathcal{F}_0)$. Since $\mathbb{P} = \mathbb{P}^*|_{\mathcal{F}}$ is a probability measure on the $\sigma$-field $\mathcal{F}$, it follows it is also a probability measure on $\sigma(\mathcal{F}_0)$. Moreover,

$$\mathbb{P}_0(A) = \mathbb{P}(A), \qquad \forall A \in \mathcal{F}_0.$$

This means $\mathbb{P}$ is an extension of $\mathbb{P}_0$ on $\sigma(\mathcal{F}_0)$.

**QED**

Def'n 1.8. $\pi$-**system**, $\lambda$-**system** of Subsets of $\Omega$

We say $\Pi \subseteq 2^\Omega$ is a $\pi$-***system*** of subsets of $\Omega$ if

$$\forall A, B \in \Pi\, [A \cap B \in \Pi]. \qquad \textit{closure under intersection}$$

We say $\Lambda \subseteq 2^\Omega$ is a $\lambda$-***system***, if

(a) $\emptyset \in \Lambda$;

(b) $A \in \Lambda$ implies $\Omega \setminus A \in \Lambda$; and *closure under complement*

(c) for any collection $\{A_n\}_{n=1}^\infty \subseteq 2^\Omega$ of disjoint subsets of $\Omega$, $\dot\bigcup_{n=1}^\infty A_n \in \Lambda$. *closure under countable disjoint union*

**Proposition 1.3.**

Let $\mathcal{F} \subseteq 2^\Omega$. Then

$$\mathcal{F} \text{ is a } \sigma\text{-field} \iff \mathcal{F} \text{ is a } \pi\text{-system and a } \lambda\text{-system}.$$

**Proof.** ($\implies$) This direction is more-or-less trivial.

($\impliedby$) It suffices to show that $\mathcal{F}$ is closed under countable union. Let $\{A_n\}_{n=1}^\infty \subseteq 2^\Omega$. Define

$$B_n = A_n \bigcap_{i=1}^{n-1} (\Omega \setminus A_i).$$

Then note that each $B_n \in \mathcal{F}$, as $\mathcal{F}$ is closed under complement (as $\mathcal{F}$ is a $\lambda$-system) and closed under intersection (as $\mathcal{F}$ is a $\pi$-system).

By definition $\{B_n\}_{n=1}^\infty$ is a collection of pairwise disjoint subsets of $\Omega$, so by the fact that $\mathcal{F}$ is a $\lambda$-system,

$$\bigcup_{n=1}^\infty A_n = \dot\bigcup_{n=1}^\infty B_n \in \mathcal{F}.$$

**QED**

**Proposition 1.4.**

Let $\Lambda \subseteq 2^\Omega$ be a $\lambda$-system. If $A, B \subseteq \Omega$ are such that $A, A \cap B \in \Lambda$, then $A \cap (\Omega \setminus B) \in \Lambda$.

**Proof.** Since $\Lambda$ is closed under complement, $\Omega \setminus A \in \Lambda$. Since $\Omega \setminus A, B$ are disjoint, it follows $(\Omega \setminus A) \cup (A \cap B) \in \Lambda$. By taking its complement

$$A \cap (\Omega \setminus B) = \Omega \setminus ((\Omega \setminus A) \cup (A \cap B)) \in \Lambda.$$

**QED**

**Theorem 1.5.** $\pi - \lambda$ Theorem

Let $\Pi$ be a $\pi$-system and let $\Lambda$ be a $\lambda$-system. If $\Pi \subseteq \Lambda$, then $\sigma(\Pi) \subseteq \Lambda$.

**Proof.** Define

$$\lambda(\Pi) = \bigcap \{\mathcal{L} \cap \Pi : \mathcal{L} \text{ is a } \lambda\text{-system containing } \Pi\}.$$

It is a routine task to show that $\lambda(\Pi)$ is also a $\lambda$-system containing $\Pi$.

For any $A \subseteq \Omega$, define

$$\mathcal{C}_A = \{B \subseteq \Omega : A \cap B \in \lambda(\Pi)\}.$$

- *Claim 1. $\mathcal{C}_A$ is a $\lambda$-system containing $\lambda(\Pi)$.*

  Proof. Let $A \in \lambda(\Pi)$ and we check three things.

  (a) $A \cap \Omega = A \in \lambda(\Pi)$, which means $\Omega \in \mathcal{C}_A$.

  (b) Given $B \in \mathcal{C}_A$, then $\lambda(\Pi)$ is a $\lambda$-system containing both $A, A \cap B$. Then we know that $A \cap (\Omega \setminus B) \in \lambda(\Pi)$. It follows $\Omega \setminus B \in \lambda(\Pi)$.

  (c) If $\{B_n\}_{n=1}^{\infty} \subseteq \mathcal{C}_A$ is a collection of disjoint sets in $\mathcal{C}_A$, then $A \cap B_1, A \cap B_2, \ldots$ are in $\lambda(\Pi)$ and are disjoint. By taking the union of $A \cap B_n$'s,

  $$\bigcup_{n=1}^{\infty} A \cap B_n = A \cap \bigcup_{n=1}^{\infty} B_n \in \lambda(\Pi).$$

  Thus $\bigcup_{n=1}^{\infty} B_n \in \lambda(\Pi)$.

  Moreover, if $A \in \Pi$, then for any $B \in \Pi$,

  $$A \cap B \in \Pi \subseteq \lambda(\Pi).$$

  Then $B \in \mathcal{C}_A$. Hence $\Pi \subseteq \mathcal{C}_A$. Since $\mathcal{C}_A$ is a $\lambda$-system, $\lambda(\Pi) \subseteq \mathcal{C}_A$.

Now assume $A \in \Pi, B \in \lambda(\Pi)$. Then $B \in \mathcal{C}_A$, so $A \cap B \in \lambda(\Pi)$. This also means $A \in \mathcal{C}_B$. Since this holds for all $A \in \Pi$, we have

$$B \in \lambda(\Pi) \implies \Pi \subseteq \mathcal{C}_B \implies \lambda(\Pi) \subseteq \mathcal{C}_B.$$

Therefore, for any $A, B \in \lambda(\Pi)$, $A \in \mathcal{C}_B$. Hence $A \cap B \in \lambda(\Pi)$, which means $\lambda(\Pi)$ is a $\pi$-system, so that it is a $\sigma$-field. As a result,

$$\Pi \subseteq \sigma(\Pi) \subseteq \lambda(\Pi) \subseteq \Lambda.$$

**QED**

**Corollary 1.5.1.**

Let $\Pi \subseteq 2^{\Omega}$ be a $\pi$-system and suppose that two probability measures $\mathbb{P}_1, \mathbb{P}_2$ agree on $\Pi$. Then they agree on $\sigma(\Pi)$.

**Proof.** Let

$$\Lambda = \{A \in \Pi : \mathbb{P}_1(A) = \mathbb{P}_2(A)\}.$$

Claim 1. *$\Lambda$ is a $\lambda$-system.*

Note that $\mathbb{P}_1(\emptyset) = 0 = \mathbb{P}_2(\emptyset)$ so that $\emptyset \in \Lambda$.

Suppose that $A \in \Lambda$. Then

$$\mathbb{P}_1(\Omega \setminus A) = 1 - \mathbb{P}_1(A) = 1 - \mathbb{P}_2(A) = \mathbb{P}_2(\Omega \setminus A),$$

so that $\Omega \setminus A \in \Lambda$.

Let $\{A_n\}_{n=1}^{\infty} \subseteq \Lambda$ be a subcollection of disjoint sets. Then

$$\mathbb{P}_1\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}_1(A_n) = \sum_{n=1}^{\infty} \mathbb{P}_2(A_n) = \mathbb{P}_2\left(\bigcup_{n=1}^{\infty} A_n\right),$$

so that $\bigcup_{n=1}^{\infty} A_n \in \Lambda$.

*(Done with Claim 1)*

So $\Lambda$ is a $\lambda$-system containing $\Pi$. By the $\pi - \lambda$ theorem, $\Lambda \supseteq \sigma(\Pi)$. Thus $\Lambda = \sigma(\Pi)$, as required.

**QED**

The uniqueness part of the theorem follows immediately from Corollary 1.5.1 and the fact that a field is a $\pi$-system.

**Example 1.8.** Lebesgue Measure on $(0,1]$

Let $\Omega = (0,1]$ and let

$$\mathcal{B}_0 = \left\{ \dot{\bigcup_{k=1}^{n}} I_k : I_1, \ldots, I_k \subseteq (0,1] \text{ are disjoint intervals} \right\}.$$

Then $\mathcal{B}_0$ is a field.

Define $\lambda : \mathcal{B}_0 \to [0,1]$ such that

$$\lambda \left( \dot{\bigcup_{k=1}^{n}} I_k \right) = \sum_{k=1}^{n} \lambda (I_k)$$

for all $\dot{\bigcup}_{k=1}^{n} I_k \in \mathcal{B}_0$, where $\lambda (I_k) = b_k - a_k$ for any interval $I_k$ with endpoints $a_k < b_k$. Then $\lambda$ is a probability premeasure on $\mathcal{B}_0$.

So by the extension theorem, there exists a unique probability measure $\overline{\lambda} : \mathcal{B}((0,1]) \to [0,1]$ on $\sigma(\mathcal{B}_0) = \mathcal{B}((0,1])$ that extends $\lambda$.

We call $\overline{\lambda}$ the ***Lebesgue measure*** on $(0,1]$.

---

Def'n 1.9. **Complete** Measure

Let $(\Omega, \mathcal{F})$ be a measurable space. We say $\mathbb{P}$ is ***complete*** probability measure on $(\Omega, \mathcal{F})$ if $\mathbb{P}$ is a probability measure with

$$\forall A \in \mathcal{F} \left[ \mathbb{P}(A) = 0 \implies \forall B \subseteq A \left[ B \in \mathcal{A} \right] \right].$$

In this case, we say $(\Omega, \mathcal{F}, \mathbb{P})$ is a ***complete*** probability space.

If $(\Omega, \mathcal{F}, \mathbb{P})$ is a complete probability space, then for $A \subseteq \Omega$, if there is $B \in \mathcal{F}$ such that

$$A \triangle B \subseteq C$$

for some $C \in \mathcal{F}$ with $\mathbb{P}(C) = 0$, then $A \in \mathcal{F}$ with $\mathbb{P}(A) = \mathbb{P}(B)$.

**Proposition 1.6.**

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Then there exists a unique complete probability space $(\Omega, \mathcal{F}', \mathcal{P}')$ such that $\mathcal{F} \subseteq \mathcal{F}'$ and $\mathbb{P}' |_{\mathcal{F}} = \mathbb{P}$.

**Proof.** Let

$$\mathcal{M} = \{ A \subseteq \Omega : A \text{ is } \mathbb{P}^*\text{-measurable} \},$$

where $\mathbb{P}^* : 2^{\Omega} \to [0,1]$ is the outer measure extending $\mathbb{P}$. Then recall that $\mathbb{P}^*$ is a probability measure on $2^{\Omega}$.

We are going to show that $\mathbb{P}^* |_{\mathcal{M}}$ is a complete measure on $(\Omega, \mathcal{M})$. So let $A \in \mathcal{M}$ be such that $\mathbb{P}^*(B) = 0$ and let $A \subseteq B$. We must show that $A$ is $\mathbb{P}^*$-measurable, so let $E \subseteq \Omega$.

Then

$$\mathbb{P}^*(E \cap A) + \mathbb{P}^*(E \cap (\Omega \setminus A)) \leq \mathbb{P}^*(B) + \mathbb{P}^*(E) = \mathbb{P}^*(E).$$

The other direction is trivial, as usual.

Then $\mathbb{P}^*(A) = 0$ follows from the monotonicity of outer measures.

**QED**

# II. Sequence of Events

## 1. Conditional Probability

Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Def'n 2.1. **Conditional Probability**
Let $A \in \mathcal{F}$ be such that $\mathbb{P}(A) > 0$. Then we define the ***conditional probability*** of $B$ given $A$, denoted as $\mathbb{P}(B|A)$, as

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}.$$

Recall the following properties of conditional probability.

**Proposition 2.1.** Chain Rule

Let $\{A_k\}_{k=1}^{n} \subseteq \mathcal{F}$. Then

$$\mathbb{P}\left(\bigcap_{k=1}^{n} A_k\right) = \prod_{k=1}^{n} \mathbb{P}\left(A_k | \bigcap_{j=1}^{k-1} A_j\right).$$

---

**Proposition 2.2.** Law of Total Probability

Suppose that $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$ is a partition of $\Omega$. Then

$$\mathbb{P}(B) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)\, \mathbb{P}(B|A_n), \qquad \forall B \in \mathcal{F}.$$

---

## 2. Limit of Events

Recall 2.2. **Limit Superior**, **Limit Inferior**, **Limit** of a Sequence of Sets
Let $(A_n)_{n=1}^{\infty}$ be a sequence of sets. Then the ***limit superior*** of $(A_n)_{n=1}^{\infty}$, denoted as $\limsup_{n\to\infty} A_n$, is defined as

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

That is,

$$w \in \limsup_{n\to\infty} A_n \iff \forall n \in \mathbb{N}\, \exists k \geq n\, [w \in A_k].$$ [1]

The ***limit inferior*** of $(A_n)_{n=1}^{\infty}$, denoted as $\liminf_{n\to\infty} A_n$, is defined as

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k.$$

That is,

$$w \in \liminf_{n\to\infty} A_n \iff \exists n \in \mathbb{N}\, \forall k \geq n\, [w \in A_k].$$ [2]

In case

$$\limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n,$$

we say $(A_n)_{n=1}^{\infty}$ has a ***limit***, denoted as $\lim_{n\to\infty} A_n$:

$$\lim_{n\to\infty} A_n = \limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n.$$

[1] We *dub* this as $w \in A_n$ *infinitely often (i.o.)*.
[2] We *dub* this as $w \in A_n$ *almost always (a.a.)*.

**Theorem 2.3.**

Let $(A_n)_{n=1}^\infty \in \mathcal{F}^{\mathbb{N}}$.

(a) We have
$$\mathbb{P}\left(\liminf_{n\to\infty} A_n\right) \leq \liminf_{n\to\infty} \mathbb{P}(A_n) \leq \limsup_{n\to\infty} \mathbb{P}(A_n) \leq \mathbb{P}\left(\limsup_{n\to\infty} A_n\right).$$

(b) If $\lim_{n\to\infty} A_n = A$, then $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$.

**Proof.**

(a) Define $B_n = \bigcap_{k=n}^\infty A_k$, $C_n = \bigcup_{k=n}^\infty A_k$ for all $n \in \mathbb{N}$. Then $(B_n)_{n=1}^\infty$ is an increasing chain with $\bigcup_{n=1}^\infty B_n = \liminf_{n\to\infty} A_n$ and $(C_n)_{n=1}^\infty$ is an decreasing chain with $\bigcap_{n=1}^\infty C_n = \limsup_{n\to\infty} A_n$. So by the continuity of probability measure,
$$\begin{aligned} \lim_{n\to\infty} \mathbb{P}(B_n) &= \mathbb{P}\left(\liminf_{n\to\infty} A_n\right) \\ \lim_{n\to\infty} \mathbb{P}(C_n) &= \mathbb{P}\left(\limsup_{n\to\infty} A_n\right) \end{aligned}.$$

Since $B_n \subseteq A_n \subseteq C_n$, we have
$$\liminf_{n\to\infty} \mathbb{P}(A_n) \geq \lim_{n\to\infty} \mathbb{P}(B_n) = \mathbb{P}\left(\liminf_{n\to\infty} A_n\right)$$
and
$$\limsup_{n\to\infty} \mathbb{P}(A_n) \leq \lim_{n\to\infty} \mathbb{P}(C_n) = \mathbb{P}\left(\limsup_{n\to\infty} A_n\right).$$

(b) This follows immediately from the definition of set limit and (a).

**QED**

## 3. Independence

Def'n 2.3. **Independent** Events

Let $A, B \in \mathcal{F}$. We say $A, B$ are ***independent*** if
$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B).$$

Moreover, we say $A_1, \ldots, A_n \in \mathcal{F}$ are ***mutually independent*** if
$$\mathbb{P}\left(\bigcap_{i\in I} A_i\right) = \prod_{i\in I} \mathbb{P}(A_i), \qquad \forall I \subseteq \{1, \ldots, n\}.$$

We say $\mathcal{A} \subseteq \mathcal{F}$ is ***independent*** if for every finite $\mathcal{B} \subseteq \mathcal{A}$,
$$\mathbb{P}\left(\bigcap_{B\in\mathcal{B}} B\right) = \prod_{B\in\mathcal{B}} \mathbb{P}(B).$$

We say $\{\mathcal{A}_\theta\}_{\theta\in\Theta} \subseteq \mathcal{P}(\mathcal{F})$ is ***independent*** if, given any $A_\theta \in \mathcal{A}_\theta$ for all $\theta \in \Theta$, $\{A_\theta\}_{\theta\in\Theta}$ is independent.

Let $A, B \in \mathcal{F}$. If $\mathbb{P}(A) > 0$, then $A, B$ are independent if and only if $\mathbb{P}(B|A) = \mathbb{P}(B)$.

Mutual independence is stronger than

(a) pairwise independence: $A_i, A_j$ are independet for all $i \neq j$; and

(b) $\mathbb{P}\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n \mathbb{P}(A_i)$.

Def'n 2.4. **Independent** $\sigma$-fields

Let $\mathcal{F}_\theta$, $\theta \in \Theta$, be $\sigma$-fields on $\Omega$. We say $\mathcal{F}_\theta$, $\theta \in \Theta$, are ***independent*** if every $A_\theta \in \mathcal{F}_\theta$, $\theta \in \Theta$, are independent.

**Proposition 2.4.**

Suppose that $\{\mathcal{A}_\theta\}_{\theta\in\Theta} \subseteq \mathcal{P}(\mathcal{F})$ is independent and suppose each $\mathcal{A}_\theta$ is a $\pi$-system. Then $\sigma(\mathcal{A}_\theta)$'s are independent.

**Proposition 2.5.**

Let

$$\begin{matrix} A_{1,1} & A_{1,2} & \cdots \\ A_{2,1} & A_{2,2} & \cdots \\ \vdots & \vdots & \ddots \end{matrix}$$

be an (infinite) array of independent events. If $\mathcal{F}_i$ is the $\sigma$-field generated by the $i$th row (i.e. $\mathcal{F}_i = \sigma\left(\left\{A_{i,j}\right\}_{j\in\mathbb{N}}\right)$), then $\mathcal{F}_1, \ldots$ are independent.

**Proof.** Let

$$\mathcal{A}_i = \left\{\bigcap_{j\in J} A_{i,j} : J \subseteq \mathbb{N}, |J| < \infty\right\},$$

the collection of all finite intersections of sets in the $i$th row. Then each $\mathcal{A}_i$ is a $\pi$-system with $\sigma(\mathcal{A}_i) = \mathcal{F}_i$. By Proposition 2.4, it remains to show that $\{\mathcal{A}_i\}_{i\in\mathbb{N}}$ are independent.

Let $I \subseteq \mathbb{N}$ be any finite set of indices. For all $i \in I$, let $C_i \in \mathcal{A}_i$. That is, there is finite $J_i \subseteq \mathbb{N}$ such that

$$C_i = \bigcap_{j\in J_i} A_{i,j}.$$

Then

$$\mathbb{P}\left(\bigcap_{i\in I} C_i\right) = \mathbb{P}\left(\bigcap_{i\in I}\bigcap_{j\in J_i} A_{i,j}\right) = \prod_{i\in I}\prod_{j\in J_i}\mathbb{P}(A_{i,j}) = \prod_{i\in I}\mathbb{P}\left(\bigcap_{j\in J_i} A_{i,j}\right) = \prod_{i\in I}\mathbb{P}(C_i).$$

Thus $\{\mathcal{A}_i\}_{i\in\mathbb{N}}$ is independent, as required.

**QED**

**Theorem 2.6.** First Borel-Cantelli Lemma

Let $(A_n)_{n=1}^\infty \in \mathcal{F}^\mathbb{N}$. If $\sum_{n=1}^\infty \mathbb{P}(A_n) < \infty$, then

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

**Proof.** Recall that $\limsup_{n\to\infty} A_n = \bigcap_{n=1}^\infty \bigcup_{k=n}^\infty A_k$.

For any $m \in \mathbb{N}$, it follows that

$$\limsup_{n\to\infty} A_n \subseteq \bigcup_{k=m}^\infty A_k.$$

Hence

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) \leq \mathbb{P}\left(\bigcup_{k=m}^\infty A_k\right) \leq \sum_{k=m}^\infty \mathbb{P}(A_k).$$

But we know $\sum_{n=1}^\infty \mathbb{P}(A_k)$ converges, so by letting $m \to \infty$, we see that

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) \leq \lim_{m\to\infty}\sum_{k=m}^\infty \mathbb{P}(A_k) = 0.$$

Thus $\mathbb{P}(\limsup_{n\to\infty} A_n) = 0$, as required.

**QED**

**Theorem 2.7.** Second Borel-Cantelli Lemma

Let $(A_n)_{n=1}^{\infty} \mathcal{F}^{\mathbb{N}}$ be independent.[1] If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, then

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 1.$$

[1] This independence condition is crucial. As an exercise, find $(A_n)_{n=1}^{\infty}$ that is not independent with $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ but $\mathbb{P}(\limsup_{n\to\infty} A_n) < 1$.

**Proof.** By definition $\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$, it suffices to prove

$$\mathbb{P}\left(\Omega \setminus \left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right)\right) = 0.$$

Note that

$$\Omega \setminus \left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} (\Omega \setminus A_k).$$

Claim 1. *For all $n \in \mathbb{N}$, $\mathbb{P}\left(\bigcap_{k=n}^{\infty} \Omega \setminus A_k\right) = 0$.*

Let $N > n$. Then

$$\mathbb{P}\left(\bigcap_{k=n}^{N} \Omega \setminus A_k\right) = \prod_{k=n}^{N} \mathbb{P}(\Omega \setminus A_k) = \prod_{k=n}^{N} (1 - \mathbb{P}(A_k)) \leq \prod_{k=n}^{N} e^{-\mathbb{P}(A_k)} = e^{-\sum_{k=n}^{N} \mathbb{P}(A_k)}$$

by using the fact that $e^{-x} \geq 1 - x$ for all $x \in \mathbb{R}$. It follows that

$$\lim_{N\to\infty} \mathbb{P}\left(\bigcap_{k=n}^{N} \Omega \setminus A_k\right) = \lim_{N\to\infty} e^{-\sum_{k=n}^{N} \mathbb{P}(A_k)} = e^{\lim_{N\to\infty} -\sum_{k=n}^{N} \mathbb{P}(A_k)} = 0,$$

since $\lim_{N\to\infty} -\sum_{k=n}^{N} \mathbb{P}(A_k) = -\infty$. But $\left(\bigcap_{k=n}^{N} \Omega \setminus A_k\right)_{N>n}^{\infty}$ is a decreasing chain, so by the continuity from above,

$$\mathbb{P}\left(\bigcap_{k=n}^{\infty} \Omega \setminus A_k\right) = 0.$$

*(End of Claim 1)*
*(End of Claim 1)*

Since countable union of null events is again null, the desired equality

$$\mathbb{P}\left(\Omega \setminus \left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right)\right) = 0$$

follows.

**QED**

**Example 2.1.**

Suppose that we have balls $1, \ldots, n$ at time $n$ and we choose a ball at each time. Given $k \in \mathbb{N}$, how many times will ball $k$ be picked in total?

**Answer.** Let $A_n$ be the event that ball $k$ picked at time $n$. Then

$$\mathbb{P}(A_n) = \begin{cases} 0 & \text{if } n < k \\ \frac{1}{n} & \text{if } n \geq k \end{cases}, \qquad \forall n \in \mathbb{N}.$$

Note that $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$. Hence by the second Borel-Cantelli lemma, $\mathbb{P}(\limsup_{n\to\infty} A_n) = 1$. Thus the ball $k$ will be picked for infinite number of times $\mathbb{P}$-almost surely.

**QED**

**Example 2.2.**

Consider the setting of Example 2.1. Instead of heaving balls $1, \ldots, n$ at time $n$, we have balls $1, \ldots, 2^n$. How many times will ball $k$ picked in total?

**Answer.** Note that

$$\mathbb{P}(A_n) = \begin{cases} 0 & \text{if } 2^n < k \\ \frac{1}{2^n} & \text{if } 2^n \geq k \end{cases}, \qquad \forall n \in \mathbb{N}.$$

This means $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, so by the first Borel-Cantelli lemma, $\mathbb{P}(\limsup_{n\to\infty} A_n) = 0$. Thus ball $k$ will be picked for finitely many times $\mathbb{P}$-almost surely.

**QED**

**Example 2.3.** DTMC

In a *discrete-time Marcov chain* (*DTMC*), if a state $i$ is *recurrent*, then the chain will visit $i$ infinitely many times almost surely, given that the chain visits $i$ at least once. If $i$ is *transient*, then visiting $i$ infinitely many times happens with probability 0.

We introduce a notion leading to 0-1 laws.

Def'n 2.5. **Tail $\sigma$-field** of Collection of Events

Let $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$. We define the ***tail $\sigma$-field*** of $\{A_n\}_{n=1}^{\infty}$, denoted as $\mathcal{T}(\{A_n\}_{n=1}^{\infty})$ (or $\mathcal{T}$ when context is clear), by

$$\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma(\{A_k\}_{k=n}^{\infty}).$$

**Example 2.4.**

Let $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$. Recall that

$$\limsup_{n\in\mathbb{N}} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

In particular, $\bigcup_{k=n}^{\infty} A_k \in \sigma(\{A_k\}_{k=n}^{\infty})$ for all $n \in \mathbb{N}$, so that $\limsup_{n\in\mathbb{N}} A_n \in \mathcal{T}$.

Similarly, $\liminf_{n\in\mathbb{N}} A_n \in \mathcal{T}$.

**Theorem 2.8.** Kolmogorov's 0-1 Law

Let $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$ be independent and let $\mathcal{T}$ be the tail $\sigma$-field of $\{A_n\}_{n=1}^{\infty}$. Then

$$\mathbb{P}(A) = 0 \text{ or } \mathbb{P}(A) = 1, \qquad \forall A \in \mathcal{T}.$$

In words, *events in the tail $\sigma$-field generated by independent events are trivial.*

**Proof.** Consider application of Proposition 2.5 to

$$\begin{matrix} A_1 & A_2 & A_3 & \cdots \\ A_2 & A_3 & A_4 & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ A_n & A_{n+1} & A_{n+2} & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{matrix}.$$

In particular we see that, for all $n \in \mathbb{N}$,

$$\sigma(A_1), \ldots, \sigma(A_{n-1}), \sigma(A_n, A_{n+1}, \ldots)$$

are independent.

For any $A \in \mathcal{T}$, $A \in \sigma(A_n, A_{n+1}, \ldots)$ for any $n \in \mathbb{N}$. Hence $A$ is independent of $A_1, \ldots, A_n$. As this holds for all $A$, $(A, A_1, A_2, \ldots)$ is an independet sequence of events. This means $\sigma(A), \sigma(A_1, A_2, \ldots)$ are independent. However, we also have

$$A \in \mathcal{T} \subseteq \sigma(A_1, A_2, \ldots),$$

and $A \in \sigma(A)$. Thus $A$ is independent of itself, so that $\mathbb{P}(A) = \mathbb{P}(A)\mathbb{P}(A)$, which happens if and only if $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.

**QED**

# III. Random Variables

Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ throughout.

## 1. Random Variables

Def'n 3.1. **Measurable** Function
Let $(X, \mathcal{A}), (Y, \mathcal{B})$ be measurable spaces. We say $f : X \to Y$ is ***measurable*** if

$$f^{-1}(B) \in \mathcal{A}, \qquad \forall B \in \mathcal{B}.$$

Def'n 3.2. **Random Variable**
We say $X : \Omega \to \mathbb{R}$ is a ***random variable*** if

$$X^{-1}(A) \in \mathcal{F}, \qquad \forall A \in \mathrm{Bor}(\mathbb{R}).$$

Whe we discuss the events defined by the value of $X$, we shall use the shorthand notation like

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

**Example 3.1.**
Let $\left(\Omega, 2^{\Omega}, \mathbb{P}\right)$ be a discrete probability space. Then any $X : \Omega \to \mathbb{R}$ is a random variable.

---

**Example 3.2.**
Given any $A \in \mathcal{F}$, the *indicator function*

$$\begin{aligned} \chi_A : \Omega &\to \mathbb{R} \\ \omega &\mapsto \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases} \end{aligned}$$

of $A$ is a random variable.

---

**Example 3.3.**
Let $X : \Omega \to \mathbb{R}$ be a random variable. Then $X$ induces a probability measure $\mu : \mathbb{R} \to \mathbb{R}$ on $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}))$ by

$$\mu(B) = \mathbb{P}(X \in B) = \mathbb{P}\left(X^{-1}(B)\right), \qquad \forall B \in \mathrm{Bor}(\mathbb{R}).$$

We call $\mu$ the ***distribution*** of $X$.

---

Def'n 3.3. **Cumulative Distribution Function** (**CDF**) of a Random Variable
Let $X : \Omega \to \mathbb{R}$ be a random variable. We define the ***cumulative distribution function*** (***cdf***) of $X$, denoted as $F_X$ (or $F$ when $X$ is understood), by

$$\begin{aligned} F_X : \mathbb{R} &\to \mathbb{R} \\ x &\mapsto \mathbb{P}(X \le x) \end{aligned}.$$

Note that $F(x) = \mu((-\infty, x])$ for all $x \in \mathbb{R}$, where $\mu$ is the distribution of $X$ introduced in Example 3.3. Since $\{(-\infty, x]\}_{x \in \mathbb{R}}$ is a $\pi$-system generating $\mathrm{Bor}(\mathbb{R})$, $F$ characterizes $\mu$.

**Proposition 3.1.** Properties of CDF

Let $X : \Omega \to \mathbb{R}$ be a random variable and let $F : \mathbb{R} \to \mathbb{R}$ be the cdf of $X$.

(a) $F$ is non-decreasing.

(b) $\lim_{x\downarrow-\infty} F(x) = 0$.

(c) $\lim_{x\uparrow\infty} F(x) = 1$.

(d) $F$ is right-continuous (i.e. $\lim_{y\downarrow x} F(y) = F(x)$ for all $x \in \mathbb{R}$).

(e) $\mathbb{P}(X < x) = \lim_{y\uparrow x} F(y)$.

(f) $\mathbb{P}(X = x) = F(x) - \lim_{y\uparrow x} F(y)$.

Moreover, the conditions (a), …, (d) characterizes $F$, That is, $G : \mathbb{R} \to \mathbb{R}$ is non-decreasing and right-continuous with

$$\lim_{x\downarrow-\infty} G(x) = 0, \lim_{x\uparrow\infty} G(x) = 1,$$

then $F = G$.

**Proof.**

(a) We have

$$\begin{aligned} x_1 \le x_2 &\implies (-\infty, x_1] \subseteq (-\infty, x_2] \\ &\implies F(x_1) = \mathbb{P}(X \in (-\infty, x_1]) \le \mathbb{P}(X \in (-\infty, x_2]) = F(x_2) \end{aligned}$$

by monotonicity of $\mathbb{P}$.

(b) Note that

$$\lim_{x\downarrow-\infty} F(x) = \lim_{x\downarrow-\infty} \mathbb{P}(X \le x) = \lim_{x\downarrow-\infty} \mathbb{P}\left(X^{-1}((-\infty, x])\right) = 0$$

by the continuity of $\mathbb{P}$ from above.

(c) Similar to (b), this follows from the continuity of $\mathbb{P}$ from below.

(d) Note that

$$\lim_{y\downarrow x} X^{-1}((-\infty, y]) = \bigcap_{y\ge x} X^{-1}((-\infty, y]) = X^{-1}((-\infty, x]),$$

so by the continuity from above,

$$\lim_{y\downarrow x} F(y) = \cdots = F(x).$$

(e) We have

$$\lim_{y\uparrow x} X^{-1}((-\infty, y]) = \bigcup_{y\le x} X^{-1}((-\infty, y]) = X^{-1}((-\infty, x)).$$

(f) This follows from (d), (e).

A more stronger result of the last statement is proven in the following theorem.

QED

**Theorem 3.2.**

Let $F : \mathbb{R} \to \mathbb{R}$ be a right-continuous, non-decreasing function with $\lim_{x\downarrow-\infty} F(x) = 0, \lim_{x\uparrow\infty} F(x) = 1$. Then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $X : \Omega \to \mathbb{R}$ such that $F$ is the cdf of $X$.

**Proof.** Take $\Omega = (0,1), \mathcal{F} = \text{Bor}((0,1))$ and let $\mathbb{P}$ be the Lebesgue measure on $(\Omega, \mathcal{F})$. Define $X : \Omega \to \mathbb{R}$ by

$$X(\omega) = \sup\{y \in \mathbb{R} : F(y) < \omega\}, \qquad \forall\omega \in \Omega.$$

Let $\omega \in (0,1)$ and let $x \in \mathbb{R}$.

Claim 1. *$\omega \le F(x)$ if and only if $X(\omega) \le x$.*

If $\omega \le F(x)$, then $x > y$ for any $y$ with $F(y) < \omega$ by monotonicity of $F$. This means $x \ge X(\omega)$.

On the other hand, if $F(x) < \omega$, then by the right-continuity of $F$, there exists $x' > x$ such that $\omega > F(x')$. This means $X(\omega) = \sup\{y \in \mathbb{R} : F(y) < \omega\} \ge x' > x$.

*(End of Claim 1)*

By Claim 1,

$$\mathbb{P}(X \le x) = \mathbb{P}((0, F(x)]) = F(x).$$

Thus $F$ is the cdf of $X$.

**QED**

Def'n 3.4. Random Variables **Equal in Distribution**

Let $X, Y : \Omega \to \mathbb{R}$ be random variables. If $F_X = F_Y$, then we say $X, Y$ are ***equal in distribution*** and we write $X \overset{\mathrm{d}}{=} Y$.

$X \overset{\mathrm{d}}{=} Y$ does not guarantee that $X = Y$.

Def'n 3.5. **Probability Density Function** (**PDF**) of a Random Variable

Let $X : \Omega \to \mathbb{R}$ be a random variable. If $f_X : \mathbb{R} \to \mathbb{R}$ is such that

$$\mathbb{P}(X \le x) = \int_{-\infty}^{x} f_X(t)\,dt, \qquad \forall x \in \mathbb{R},$$

then we say $f_X$ the ***probability density function*** (***pdf***) of $X$.

Note that, if $X$ has a pdf $f_X : \mathbb{R} \to \mathbb{R}$, then

$$\mathbb{P}(X \in (a,b]) = \mathbb{P}((-\infty, b]) - \mathbb{P}((-\infty, a]) = \int_{-\infty}^{b} f_X(t)\,dt - \int_{-\infty}^{a} f_X(t)\,dt = \int_{a}^{b} f_X(t)\,dt.$$

Moreover,

$$\mathbb{P}(X = x) \le \mathbb{P}(X \in (x-\varepsilon, x+\varepsilon]) = \int_{x-\varepsilon}^{x+\varepsilon} f_X(t)\,dt$$

for any $\varepsilon > 0$, so by taking $\varepsilon \to 0$, we see that

$$\mathbb{P}(X = x) = 0.$$

Def'n 3.6. **Continuous**, **Absolutely Continuous** Random Variable

Let $X : \Omega \to \mathbb{R}$ be a random variable. If the pdf $F_X$ of $X$ is continuous, then we say $X$ is ***continuous***.

In case $X$ admits a density function $f_X$, we say $X$ (and $F_X$) is ***absolutely continuous***.

We see that an absolutely continuous random variable is continuous, but the converse is false.

Def'n 3.7. **Singular** Continuous Random Variable

We say a continuous random variable $X$ is ***singular*** if it does not admit a pdf.

Hence in general, a cdf consists of three parts: absolutely continuous, continuous, and discrete.

## 2. Examples of Distributions

Given a random variable $X$ and a distribution $F$, we write $X \sim F$ if $F$ is the cdf of $X$.

**Example 3.4.** Uniform Distribution on $(0,1)$

If $X \sim \mathrm{U}(0,1)$, then

$$f_X(x) = \begin{cases} 0 & \text{if } x \notin (0,1) \\ 1 & \text{if } x \in (0,1) \end{cases}, \qquad \forall x \in \mathbb{R}$$

and

$$F_X(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{if } x \in (0,1) \\ 1 & \text{if } x \geq 1 \end{cases}, \qquad \forall x \in \mathbb{R}.$$

---

**Example 3.5.** Exponential Distribution

If $X \sim \mathrm{EXP}(\lambda)$, where $\lambda \geq 0$, then

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}, \qquad \forall x \in \mathbb{R}.$$

---

**Proposition 3.3.**

Let $X : \Omega \to \mathbb{R}$ be a random variable. Then

$$\mathcal{G} = \left\{ X^{-1}(B) : B \in \mathrm{Bor}(\mathbb{R}) \right\}$$

is the smallest $\sigma$-field on $\Omega$ such that $X$ is measurable on $(\Omega, \mathcal{G})$.

**Proof.** Note that $\emptyset = X^{-1}(\emptyset)$ where $\emptyset \in \mathrm{Bor}(\mathbb{R})$.

Given $A = X^{-1}(B)$ for some $B \in \mathrm{Bor}(\mathbb{R})$, $\Omega \setminus A = X^{-1}(\mathbb{R} \setminus B) \in \mathcal{G}$ as $\mathbb{R} \setminus B \in \mathrm{Bor}(\mathbb{R})$.

Let $A_1 = X^{-1}(B_1), A_2 = X^{-1}(B_2), \ldots \in \mathcal{G}$ for some $B_1, B_2, \ldots \in \mathrm{Bor}(\mathbb{R})$. Then

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} X^{-1}(B_n) = X^{-1}\left( \bigcup_{n=1}^{\infty} B_n \right),$$

where $\bigcup_{n=1}^{\infty} B_n \in \mathrm{Bor}(\mathbb{R})$, so that $\bigcup_{n=1}^{\infty} A_n \in \mathcal{G}$.

Hence $\mathcal{G}$ is a $\sigma$-field.

Note that if $\mathcal{E} \subset \mathcal{G}$ is a $\sigma$-field properly contained in $\mathcal{G}$, then there is $B \in \mathrm{Bor}(\mathbb{R})$ such that $X^{-1}(B) \notin \mathcal{E}$. Hence $X$ is not measurable on $(X, \mathcal{E})$.

**QED**

Def'n 3.8. $\sigma$-field **Generated** by a Random Variable

Consider the setting of Proposition 3.3. We call $\mathcal{G}$ the $\sigma$-field ***generated*** by $X$.

**Theorem 3.4.**

Let $X : (\Omega, \mathcal{F}) \to (S, \mathcal{A})$ and $f : (S, \mathcal{A}) \to (T, \mathcal{B})$ be measurable, then $f \circ X$ is measurable.

**Proof.** Let $B \in \mathcal{B}$. Then

$$(f \circ X)^{-1}(B) = X^{-1}\left(f^{-1}(B)\right) \in \mathcal{F}$$

by the measurability of $f, X$.

**QED**

In particular, in case $X$ is a random variable, then $f \circ X$ is also a random variable.

**Theorem 3.5.**

Let $X_1, X_2, \ldots$ be random variables. Then $\inf_{n \in \mathbb{N}} X_n$, $\sup_{n \in \mathbb{N}} X_n$, $\liminf_{n \to \infty} X_n$, $\limsup_{n \to \infty} X_n$ are random variables.

**Proof.** Note that

$$\left(\inf_{n\in\mathbb{N}} X_n\right)^{-1}((-\infty,a)) = \bigcup_{n\in\mathbb{N}} X_n^{-1}((-\infty,a)) \in \mathcal{F}.$$

Similarly,

$$\left(\sup_{n\in\mathbb{N}} X_n\right)^{-1}((a,\infty)) = \bigcup_{n\in\mathbb{N}} X_n^{-1}((a,\infty)) \in \mathcal{F}.$$

Then it remains to note that $\liminf_{n\to\infty} X_n = \sup_{n\in\mathbb{N}} \inf_{k\geq n} X_k$ and that $\limsup_{n\to\infty} X_n = \inf_{n\in\mathbb{N}} \sup_{k\geq n} X_k$.

**QED**

In case $\liminf_{n\to\infty} X_n = \limsup_{n\to\infty} X_n$, $\lim_{n\to\infty} X_n$ exists and is also a random variable.

Moreover,

$$A = \left\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) \text{ exists}\right\} = \left(\limsup_{n\to\infty} X_n - \liminf_{n\to\infty} X_n\right)^{-1}(\{0\})$$

is measurable. Hence it follows that

$$\mathbb{P}(A) = 1 \iff X_n \text{ converges almost surely.}$$

# IV. Lebesgue Integration

Fix a $\sigma$-finite measure space $(\Omega, \mathcal{F}, \mu)$.

## 1. Simple Functions

Def'n 4.1. **Simple** Function
Let $\varphi : \Omega \to \mathbb{R}$. If

$$\varphi = \sum_{i=1}^{n} a_i \chi_{A_i}$$

for some $a_1, \ldots, a_n \in \mathbb{R}$ and $A_1, \ldots, A_n \in \mathcal{F}$ with each $\mu(A_i) < \infty$, then we say $\varphi$ is a ***simple*** function.

In this case, we define the ***integral*** of $\varphi$, denoted as $\int \varphi d\mu$, as

$$\int \varphi d\mu = \sum_{i=1}^{n} a_i \mu(A_i).$$

**Lemma 4.1.**
Let $\varphi : \Omega \to \mathbb{R}$ be simple with $\varphi \geq 0$ $\mu$-ae. Then

$$\int \varphi d\mu \geq 0.$$

**Lemma 4.2.** Linearity of Integration for Simple Functions
Let $\varphi, \psi : \Omega \to \mathbb{R}$ be simple functions and let $a \in \mathbb{R}$. Then

$$\int a\varphi + \psi d\mu = a \int \varphi d\mu + \int \varphi d\mu.$$

**Proof.** Write $\varphi = \sum_{i=1}^{n} a_i \chi_{A_i}$, $\psi = \sum_{j=1}^{m} b_j \chi_{B_j}$. Then note that

$$a\varphi + \psi = \sum_{i=1}^{n} \sum_{j=1}^{m} (aa_i + b_j) \chi_{A_i \cap B_j},$$

so that

$$\begin{aligned}
\int a\varphi + \psi d\mu &= \sum_{i=1}^{n} \sum_{j=1}^{m} (aa_i + b_j) \mu(A_i \cap B_j) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} aa_i \mu(A_i \cap B_j) + \sum_{i=1}^{n} \sum_{j=1}^{m} b_j \mu(A_i \cap B_j) \\
&= \sum_{i=1}^{n} aa_i \sum_{j=1}^{m} \mu(A_i \cap B_j) + \sum_{j=1}^{m} b_j \sum_{i=1}^{n} \mu(A_i \cap B_j) \\
&= a \sum_{i=1}^{n} a_i \mu(A_i) + \sum_{j=1}^{m} b_j \mu(B_j) \\
&= a \int \varphi d\mu + \int \psi d\mu.
\end{aligned}$$

**QED**

**Lemma 4.3.**

Let $\varphi, \psi : \Omega \to \mathbb{R}$.

(a) If $\varphi \leq \psi$ $\mu$-ae, then $\int \varphi d\mu \leq \int \psi d\mu$.

(b) If $\varphi = \psi$ $\mu$-ae, then $\int \varphi d\mu = \int \psi d\mu$.

(c) $\left|\int \varphi d\mu\right| \leq \int |\varphi| \, d\mu$.

**Proof.**

(a) We have

$$\int \varphi d\mu = \int \psi d\mu + \int (\varphi - \psi) \, d\mu \geq \int \psi d\mu$$

by Lemma 4.1.

(b) We note that

$$\varphi = \psi \ \mu\text{-ae} \implies \varphi \leq \psi \ \mu\text{-ae and } \psi \leq \varphi \ \mu\text{-ae}.$$

Hence by (a) $\int \varphi d\mu = \int \psi d\mu$.

(c) Note that

$$\int \varphi d\mu \leq \int |\varphi| \, d\mu$$

and that

$$\int -\varphi d\mu \leq \int |\varphi| \, d\mu$$

so that

$$|\varphi d\mu| \leq \int |\varphi| \, d\mu.$$

**QED**

## 2. Integration of Bounded Functions on a Finite Measure Support

**Proposition 4.4.**

Let $f : \Omega \to \mathbb{R}$ be a bounded function such that there is $E \in \mathcal{F}$ with $\mu(E) < \infty$ such that $f(x) = 0$ for all $x \in \Omega \setminus E$. Then

$$\underbrace{\sup\left\{\int \varphi d\mu : \varphi \text{ is simple and } \varphi \leq f\right\}}_{=l} = \underbrace{\inf\left\{\int \psi d\mu : \psi \text{ is simple and } \psi \geq f\right\}}_{=u}.$$

**Proof.** By monotonicity, $l \leq u$ is clear.

We are going to show that the difference $u - l$ is dominated by a positive sequence converging to 0. Let $M \in \mathbb{R}$ be such that $|f(x)| \leq M$ for all $x \in E$ and let

$$E_{k,n} = \left\{x \in E : \frac{(k-1)\,M}{n} < f(x) \leq \frac{kM}{n}\right\}, \qquad \forall n \in \mathbb{N}, k \in \{-n, \ldots, n\}.$$

Then by taking

$$\psi_n = \sum_{k=-n}^{n} \frac{kM}{n} \chi_{E_{k,n}}, \varphi_n = \sum_{k=-n}^{n} \frac{(k-1)\,M}{n} \chi_{E_{k,n}}, \qquad \forall k \in \{-n, \ldots, n\},$$

we see that $\varphi_n \leq f \leq \psi_n$ and that

$$\int \psi_n d\mu - \int \varphi_n d\mu = \int \psi_n - \varphi_n d\mu = \frac{M}{n} \mu(E) \overset{n \to \infty}{\to} 0.$$

**QED**

**Lemma 4.5.**

Lemma 4.1, 4.2, 4.3 applies for bounded function with finite measure support.

## 3. Integration of Nonnegative Measurable Function

Def'n 4.2. **Integral** of Nonnegative Measurable Function

Let $f : X \to \mathbb{R}$ be a measurable function such that $f \geq 0$. We define the ***integral*** of $f$, denoted as $\int fd\mu$, by

$$\int fd\mu = \sup \left\{ \int hd\mu : 0 \leq h \leq f, h \text{ is bounded}, \mu \left( \{x \in X : h(x) \neq 0\} \right) < \infty \right\}.$$

A way to find $\int fd\mu$ is to consider

$$h_n = (f \wedge n) \chi_{E_n}, \qquad \forall n \in \mathbb{N},$$

where $(E_n)_{n=1}^{\infty} \in \mathcal{A}^{\mathbb{N}}$ is an increasing chain of sets in $\mathcal{A}$ with $\mu(E_n) < \infty$ such that $\bigcup_{n=1}^{\infty} E_n = \Omega$, which exists by the fact that $(X, \mathcal{A}, \mu)$ is $\sigma$-finite.

**Lemma 4.6.**

Let $f : X \to \mathbb{R}$ be nonnegative and let

$$h_n = (f \wedge n) \chi_{E_n}, \qquad \forall n \in \mathbb{N},$$

where $(E_n)_{n=1}^{\infty} \in \mathcal{A}^{\mathbb{N}}$ is an increasing chain of sets in $\mathcal{A}$ with $\mu(E_n) < \infty$ such that $\bigcup_{n=1}^{\infty} E_n = \Omega$. Then

$$\int h_n d\mu \overset{n\to\infty}{\nearrow} \int fd\mu.$$

**Proof.** Note that $\left(\int h_n d\mu\right)_{n=1}^{\infty}$ is an increasing sequence of nonnegative numbers, so

$$\lim_{n\to\infty} \int h_n d\mu$$

exists (in $[0, \infty]$). Let $h : X \to \mathbb{R}$ be bounded with $0 \leq h \leq f$ and a finite measure support. Let $M \in \mathbb{R}$ be such that $h(x) \leq M$ for all $x \in X$. Then for every $n \geq M$,

$$\int h_n d\mu = \int_{E_n} f \wedge n d\mu \geq \int_{E_n} hd\mu = \int hd\mu - \int_{X \setminus E_n} hd\mu$$

But note that

$$\int_{X \setminus E_n} hd\mu = \int_{(X \setminus E_n) \cap E} hd\mu \leq \int_{(X \setminus E_n) \cap E} Md\mu = M\mu\left((X \setminus E_n) \cap E\right) = M\mu(E \setminus E_n) \overset{n\to\infty}{\to} 0.$$

This means

$$\lim_{n\to\infty} \int h_n d\mu \geq \int hd\mu.$$

By taking sup over $h$, we see that

$$\lim_{n\to\infty} h_n d\mu \geq \int fd\mu.$$

The reverse inequality

$$\lim_{n\to\infty} \int h_n d\mu \leq \int fd\mu$$

is clear from the fact that each $h_n$ is bounded with $0 \leq h_n \leq f$ and $\mu\left(h_n^{-1}(\mathbb{R} \setminus \{0\})\right) \leq \mu(E_n) < \infty$, so that

$$\int h_n d\mu \leq \int fd\mu.$$

**QED**

**Lemma 4.7.**

Lemma 4.1, 4.2, 4.3 holds for nonnegative measurable functions.

## 4. Integrable Functions

Def'n 4.3. **Integrable** Function

Let $f : X \to \mathbb{R}$ be measurable. We say $f$ is ***integrable*** if $\int |f|\, d\mu < \infty$.

When $f$ is integrable, we define the ***positive part*** $f^+$ and ***negative part*** $f^-$ of $f$ by

$$f^+ = f \vee 0, f^- = -(f \wedge 0).$$

We note that $f = f^+ - f^-$ and that $|f| = f^+ + f^-$.

Def'n 4.4. **Integral** of an Integrable Function

Let $f : X \to \mathbb{R}$ be integrable. We define the ***integral*** of $f$, denoted as $\int f d\mu$, by

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

**Theorem 4.8.**

Lemma 4.1, 4.2, 4.3 hold for integrable functions.

## 5. Limit Theorems of Integration

We present few useful limit theorems without proof.

**Theorem 4.9.** Fatou's Lemma

Let $(f_n)_{n=1}^{\infty}$ be a sequence of nonnegative integrable function. Then

$$\liminf_{n\to\infty} \int f_n d\mu \geq \int \left(\liminf_{n\to\infty} f_n\right) d\mu,$$

where $\liminf_{n\to\infty} f_n$ is taken pointwise.

**Theorem 4.10.** Monotone Convergence Theorem (MCT)

Let $(f_n)_{n=1}^{\infty}$ be an increasing sequence of nonnegative integrable functions and let $f : X \to [0, \infty]$ by

$$f(x) = \lim_{n\to\infty} f_n(x), \qquad \forall x \in X.$$

Then

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

**Theorem 4.11.** Dominated Convergence Theorem (DCT)

Let $(f_n)_{n=1}^{\infty}$ be a sequence of integrable function converging pointwise to $f : X \to \mathbb{R}$ $\mu$-ae and suppose $|f_n| \leq g$ for some integrable $g : X \to \mathbb{R}$. Then

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

# V. Order Statistics

Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and fix a random variable $X$ on $(\Omega, \mathcal{F}, \mathbb{P})$.

## 1. Expectation

Def'n 5.1. **Expectation** of a Random Variable
We define the ***expectation*** of $X$, denoted as $\mathbb{E}(X)$, by

$$\mathbb{E}(X) = \int X d\,\mathbb{P}.$$

By the definition of integration,

$$\mathbb{E}(X) = \mathbb{E}\left(X^{+}\right) - \mathbb{E}\left(X^{-}\right).$$

Using this as the motivation, we define $\mathbb{E}(X) = \infty$ in case $\mathbb{E}(X^{+}) = \infty, \mathbb{E}(X^{-}) < \infty$ and $\mathbb{E}(X) = -\infty$ in case $\mathbb{E}(X^{+}) < \infty, \mathbb{E}(X^{-}) = \infty$.

In case $\mathbb{E}(X^{+}) = \mathbb{E}(X^{-}) = \infty$, we leave $\mathbb{E}(X)$ as undefined.

**Proposition 5.1.** Linearity of Expectation
Let $X, Y : \Omega \to \mathbb{R}$ be random variables and let $a \in \mathbb{R}$. Then

$$\mathbb{E}(aX + Y) = a\,\mathbb{E}(X) + \mathbb{E}(Y).$$

**Proposition 5.2.** Monotonicity of Expectation
Let $X, Y : \Omega \to \mathbb{R}$ be random variables with $X \geq Y$ almost surely. Then

$$\mathbb{E}(X) \geq \mathbb{E}(Y).$$

**Theorem 5.3.** Change of Variable
Let $X$ be a random variable with distribution $\mu$ and let $f$ be a measurable function such that

$$f \geq 0 \text{ or } \mathbb{E}(|f(X)|) < \infty.$$

Then

$$\mathbb{E}(f(X)) = \int f d\mu.$$

**Proof.** Case 1. *Consider $f = \chi_B$ for some measurable $B$.*

We have

$$\mathbb{E}(f(X)) = \mathbb{E}\left(\chi_B(X)\right) = \mathbb{P}(X \in B) = \mu(B) = \int \chi_B d\mu = \int f d\mu.$$

*(End of Case 1)*

Case 2. *Suppose f is a simple function, say*

$$f = \sum_{k=1}^{n} c_k \chi_{B_k}$$

*for some* $c_1, \ldots, c_n \in \mathbb{R}$ *and measurable* $B_1, \ldots, B_n$.

Then

$$\mathbb{E}\left(f(X)\right) = \mathbb{E}\left(\sum_{k=1}^{n} c_k \chi_{B_k}(X)\right) = \sum_{k=1}^{n} c_k \, \mathbb{E}\left(\chi_{B_k}(X)\right) = \sum_{k=1}^{n} c_k \int \chi_{B_k} d\mu = \int f d\mu.$$

*(End of Case 2)*

Case 3. *Suppose f is a nonnegative μ-measurable function.*

Let $(f_n)_{n=1}^{\infty}$ be a sequence of simple functions such that $f_n \nearrow f$ pointwise. Then by the MCT,

$$\mathbb{E}\left(f(X)\right) = \lim_{n\to\infty} \mathbb{E}\left(f_n(X)\right) = \lim_{n\to\infty} \int f_n d\mu = \int f d\mu.$$

*(End of Case 3)*

Case 4. *Suppose f is μ-integrable.*

Then $f = f^+ - f^-$, so that

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

*(End of Case 4)*

**QED**

**Corollary 5.3.1.**

Let $X$ be a random variable with density of $f$. Then

$$\mathbb{E}\left(g(X)\right) = \int g(x) f(x)\, dx.$$

## 2. Moments and Variation

Def'n 5.2. ***k*th Moment**, **Variation** of a Random Variable

Let $X$ be a random variable. We call $\mathbb{E}\left(X^k\right)$ the ***kth moment*** of $X$.

We define the ***variation*** of $X$, denoted as $\mathrm{var}(X)$, by

$$\mathrm{var}(X) = \mathbb{E}\left(\left(X - \mathbb{E}(X)\right)^2\right).$$

Note that

$$\mathrm{var}(X) = \mathbb{E}\left(\left(X - \mathbb{E}(X)\right)^2\right) = \cdots = \mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2.$$

**Example 5.1.** Bernoulli Distribution

We say a random variable $X$ is ***Bernoulli*** with probability $p \in [0,1]$ if

$$\mathbb{P}(X=1) = p, \mathbb{P}(X=0) = 1-p.$$

Hence

$$\mathbb{E}(X) = p$$

and

$$\operatorname{var}(X) = p - p^2 = p(1-p),$$

by using the fact that $X^2 = X$.

---

**Example 5.2.** Poisson Distribution

We say a random variable $X$ is ***Poisson*** with parameter $\lambda > 0$, written as $X \sim \mathrm{POI}(\lambda)$, if

$$\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad \forall k \in \mathbb{N} \cup \{0\}.$$

Note that

$$\begin{aligned}
\mathbb{E}(X(X-1)\cdots(X-k+1)) &= \sum_{j=0}^{\infty} j(j-1)\cdots(j-k+1)\, e^{-\lambda}\frac{\lambda^j}{j!} \\
&= \sum_{j=k}^{\infty} j(j-1)\cdots(j-k+1)\, e^{-\lambda}\frac{\lambda^j}{j!} \\
&= \sum_{j=k}^{\infty} e^{-\lambda}\frac{\lambda^j}{(j-k)!} \\
&= \sum_{n=0}^{\infty} e^{-\lambda}\frac{\lambda^{k+n}}{n!} \\
&= \lambda^k \sum_{n=0}^{\infty} \underbrace{e^{-\lambda}\frac{\lambda^n}{n!}}_{\text{pmf of POI}(\lambda)} \\
&= \lambda^k.
\end{aligned}$$

Consequently,

$$\mathbb{E}(X) = \lambda, \mathbb{E}(X(X-1)) = \lambda^2,$$

so that

$$\mathbb{E}\left(X^2\right) = \mathbb{E}(X(X-1)) + \mathbb{E}(X) = \lambda^2 + \lambda.$$

Thus

$$\operatorname{var}(X) = \mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2 = \lambda.$$

---

**Example 5.3.** Exponential Distribution

We say a random variable $X$ is ***Exponential*** with parameter $\lambda > 0$, written as $X \sim \mathrm{EXP}(\lambda)$, if

$$f(x) = \lambda e^{-\lambda x}, \qquad \forall x \geq 0,$$

is the pdf of $X$. Then

$$\mathbb{E}\left(X^k\right) = \int_0^{\infty} x^k \lambda e^{-\lambda x} dx = \frac{1}{\lambda^k}\int_0^{\infty} y^k e^{-y} dy = \frac{1}{\lambda^k}\Gamma(k+1) = \frac{1}{\lambda^k}k!.$$

In particular,

$$\mathbb{E}(X) = \frac{1}{\lambda}$$

and

$$\operatorname{var}(X) = \frac{1}{\lambda^2}.$$

---

# VI. Probability Inequalities

Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

## 1. Chebyshev's Inequality

**Theorem 6.1.** Generalized Chebyshev's Inequality

Let $X$ be a random variable and let $g$ be a nonnegative function. Let $B \in \mathcal{F}$ and let

$$l = \inf \{g(x) : x \in B\}.$$

Then

$$l \, \mathbb{P}(X \in B) \leq \mathbb{E}(g(X)).$$

**Proof.** Define

$$Y = l \chi_B(X).$$

Then $Y \leq l \leq g(X)$, so that

$$l \, \mathbb{P}(X \in B) = \mathbb{E}(Y) \leq \mathbb{E}(g(X)).$$

**QED**

**Corollary 6.1.1.** Markov's Inequality

In particular, if $X$ is a nonnegative random variable, then

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}, \qquad \forall a > 0.$$

**Corollary 6.1.2.**

Let $X$ be a random varaible. Then

$$a^2 \, \mathbb{P}(|X| \geq a) \leq \mathbb{E}\left(X^2\right)$$

and

$$a^2 \, \mathbb{P}(|X - \mathbb{E}(X)| \geq a) \leq \operatorname{var}(X).$$

# VII. Several Random Variables

## 1. Independence

Def'n 7.1. **Independent** Random Variables
We say two random variables $X, Y$ are ***independent*** if $\sigma(X)$ and $\sigma(Y)$ are independent.

That is, for every Borel $A, B$, the preimages $X^{-1}(A)$ and $Y^{-1}(B)$ are independent:

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

More generally, random variables $X_1, \ldots, X_n$ are ***independent*** if $\sigma(X_1), \ldots, \sigma(X_n)$ are independent. That is,

$$\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \prod_{k=1}^{n} \mathbb{P}(X_k \in A_k), \qquad \forall A_1, \ldots, A_n \in \mathrm{Bor}(\mathbb{R}).$$

**Proposition 7.1.**
Let $X, Y$ be random variables and let $f, g$ be measurable. If $X, Y$ are independent, then so are $f(X), g(Y)$.

**Proof.** For all $A, B \in \mathrm{Bor}(\mathbb{R})$,

$$\mathbb{P}(f(X) \in A, g(Y) \in B) = \mathbb{P}\left(X \in f^{-1}(A), Y \in g^{-1}(B)\right) = \mathbb{P}\left(X \in f^{-1}(A)\right)\mathbb{P}\left(Y \in g^{-1}(B)\right) = \mathbb{P}(f(X))\,\mathbb{P}(g(Y)).$$

**QED**

**Theorem 7.2.**
Let $\mathcal{A}_1, \mathcal{A}_2$ be independent $\pi$-systems. Then so are $\sigma(\mathcal{A}_1), \sigma(\mathcal{A}_2)$.

**Proof.** Let $B_2 \in \mathcal{A}_2$. Define

$$\mathcal{L} = \{B \in \mathcal{A}_1 : \mathbb{P}(B \cap B_2) = \mathbb{P}(B)\,\mathbb{P}(B_2)\}.$$

Claim 1. *$\mathcal{L}$ is a $\lambda$-system.*

Observe that

$$\mathbb{P}(\emptyset \cap B_2) = \mathbb{P}(\emptyset) = \mathbb{P}(B_2)\,\mathbb{P}(\emptyset).$$

If $B \in \mathcal{L}$, then

$$\mathbb{P}((\Omega \setminus B) \cap B_2) = \cdots = \mathbb{P}(\Omega \setminus B)\,\mathbb{P}(B_2)$$

so that $\Omega \setminus B \in \mathcal{L}$.

Lastly, if $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{L}$ is a collection of disjoint sets in $\mathcal{L}$, then

$$\mathbb{P}\left(\dot{\bigcup_{n=1}^{\infty}} A_n \cap B_2\right) = \cdots = \mathbb{P}\left(\dot{\bigcup_{n=1}^{\infty}} A_n\right)\mathbb{P}(B_2),$$

so that $\dot{\bigcup}_{n=1}^{\infty} A_n \in \mathcal{L}$.

*(End of Claim 1)*

It is immediate that

$$\mathcal{L} \supseteq \mathcal{A}_1,$$

so that

$$\sigma(\mathcal{L}) \supseteq \sigma(\mathcal{A}_1).$$

Since the above holds for any arbitrary choice of $B_2 \in \mathcal{A}_2$, it follows that any $B_1 \in \sigma(\mathcal{A}_1)$ is independent of any $B_2 \in \mathcal{A}_2$. By symmetry, any $B_2 \in \sigma(\mathcal{A}_2)$ is independent of any $B_1 \in \mathcal{A}_1$, as required.

**QED**

Def'n 7.2. **Joint CDF** of Two Random Variables
Let $X$, $Y$ be random variables. We define the ***joint cdf*** of $X$, $Y$, denoted as $F_{X,Y}$, by

$$F_{X,Y} = \mathbb{P}\left(X \leq x, Y \leq y\right), \qquad \forall x, y \in \mathbb{R} \cup \{-\infty, \infty\}.$$

Let $\eta$ be the joint distribution of $(X, Y)$. That is,

$$\eta\left(B\right) = \mathbb{P}\left(\left(X, Y\right) \in B\right), \qquad \forall B \in \mathcal{B}\left(\mathbb{R}^2\right).$$

Then

$$F_{X,Y}\left(x, y\right) = \eta\left(\left(-\infty, x\right] \times \left(-\infty, y\right]\right), \qquad \forall x, y \in \mathbb{R}.$$

Then

$$\begin{aligned} X, Y \text{ are independent } &\iff F_{X,Y}\left(x, y\right) = F_X\left(x\right) F_Y\left(y\right), \qquad \forall x, y \in \mathbb{R} \\ &\iff \mathbb{P}\left(X \leq x, Y \leq y\right) = \mathbb{P}\left(X \leq x\right) \mathbb{P}\left(Y \leq y\right). \end{aligned}$$

Proof. ( $\Longleftarrow$ ) This direction is trivial.
( $\Longrightarrow$ ) Note that $\{\{X \leq x\}\}_{x \in \mathbb{R}}, \{\{Y \leq y\}\}_{y \in \mathbb{R}}$ are $\pi$-systems that are independent. Hence

$$\sigma\left(X\right) = \sigma\left(\{\{X \leq x\}\}_{x \in \mathbb{R}}\right)$$

and

$$\sigma\left(Y\right) = \sigma\left(\{\{Y \leq y\}\}_{y \in \mathbb{R}}\right)$$

are independent. *QED*

The same result can be shown in a more general setting:

$$X_1, \ldots, X_n \text{ are independent } \iff \mathbb{P}\left(X_1 \leq x_1, \ldots, X_n \leq x_n\right) = \prod_{i=1}^{n} \mathbb{P}\left(X_i \leq x_i\right), \qquad \forall x_1, \ldots, x_n \in \mathbb{R}.$$

If we let $\mu$, $\nu$ be the distributions of $X$, $Y$, respectively, then observe that

$$\eta\left(\left(-\infty, x\right] \times \left(-\infty, y\right]\right) = \mu\left(\left(-\infty, x\right]\right) \eta\left(\left(-\infty, y\right]\right), \qquad \forall x, y \in \mathbb{R}.$$

It follows that

$$\eta\left(A \times B\right) = \mu\left(A\right) \nu\left(B\right), \qquad \forall A, B \in \mathrm{Bor}\left(\mathbb{R}\right).$$

This implies that if $X$, $Y$ are independent, then

$$\mathbb{E}\left(h\left(X, Y\right)\right) = \int h d\eta = \int h d\left(\mu \times \nu\right) = \iint h d\mu d\nu$$

for any measurable $h : \mathbb{R}^2 \to \mathbb{R}$ with $h \geq 0$ or $\mathbb{E}\left(\left|h\left(X, Y\right)\right|\right) < \infty$.

**Proposition 7.3.**
Let $X_1, \ldots, X_n$ be independent random variables. If

$$X_i \geq 0, \qquad \forall i \in [1, \ldots, n]$$

or

$$\mathbb{E}\left(\left|X_i\right|\right) < \infty, \qquad \forall i \in \{1, \ldots, n\},$$

then

$$\mathbb{E}\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} \mathbb{E}\left(X_i\right).$$

**Proof.** We consider the case $n = 2$.

By defining

$$\begin{aligned} h : \mathbb{R}^2 &\to \mathbb{R} \\ (x, y) &\mapsto xy \end{aligned},$$

note that $XY = h(X, Y)$. Hence

$$\mathbb{E}(h(X, Y)) = \int h d\eta = \iint h d\mu d\nu = \int x\mu(dx) \int y\nu(dy) = \mathbb{E}(X)\mathbb{E}(Y).$$

**QED**

## 3. Covariance and Correlation

Def'n 7.3. **Covariance** of Two Random Variables

Let $X, Y$ be random variables. We define the ***covariance*** of $X, Y$, denoted as $\operatorname{cov}(X, Y)$, by

$$\operatorname{cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$$

provided that the related quantities are well-defined.

Note that

$$\operatorname{cov}(X, Y) = \cdots = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

Def'n 7.4. **Correlation** of Two Random Variables

Let $X, Y$ be random variables. We define the ***correlation*** of $X, Y$, denoted as $\operatorname{cor}(X, Y)$, by

$$\operatorname{cor}(X, Y) = \frac{\operatorname{cov}(X, Y)}{\sqrt{\operatorname{var}(X)\operatorname{var}(Y)}}.$$

We say $X, Y$ are ***uncorrelated*** if $\operatorname{cor}(X, Y) = 0$.

If $X, Y$ are independent random variables, then

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y) \implies \operatorname{cov}(X, Y) = 0.$$

Hence

$$X, Y \text{ are independent} \implies X, Y \text{ are uncorrelated}$$

provided that the related quantities are well-defined.

That is, if $X, Y$ or $XY$ does not have finite expectation, then $\operatorname{cov}(X, Y)$ is not defined, so we cannot talk about correlation of $X, Y$.

The other direction does not hold (i.e. uncorrelatedness does not imply independence).

**Example 7.1.**

Consider $X \sim \mathcal{N}(0, 1)$, $Y = |X|$. Then $\operatorname{cor}(X, Y) = 0$ but $X, Y$ are not independent.

**Theorem 7.4.**

Let $X, Y$ be random variables.

(a) $\operatorname{var}(X + Y) = \operatorname{var}(X) + \operatorname{var}(Y) + 2\operatorname{cov}(X, Y)$.

(b) If $X, Y$ are independent, then $\operatorname{var}(X + Y) = \operatorname{var}(X) + \operatorname{var}(Y)$.

(c) $|\operatorname{cov}(X, Y)| \le \sqrt{\operatorname{var}(X)\operatorname{var}(Y)}$. Consequently, $\operatorname{cor}(X, Y) \in [-1, 1]$.

**Proof.**

(a) Note that

$$\begin{aligned}\operatorname{var}(X+Y) &= \mathbb{E}\left(\left((X+Y)-\mathbb{E}(X+Y)\right)^2\right) = \mathbb{E}\left(\left(X-\mathbb{E}(X)+(Y-\mathbb{E}(Y))\right)^2\right)\\ &= \mathbb{E}\left((X-\mathbb{E}(X))^2\right)+\mathbb{E}\left((Y-\mathbb{E}(Y))^2\right)+2\,\mathbb{E}\left((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))\right) = \operatorname{var}(X)+\operatorname{var}(Y)+2\operatorname{cov}(X,Y).\end{aligned}$$

(b) Note that $\operatorname{cov}(X,Y)=0$.

(c) Use Holder's inequality. An alternative proof can be given as follows. Note that

$$\operatorname{var}(aX-Y) = a^2\operatorname{var}(X) - 2a\operatorname{cov}(X,Y)+\operatorname{var}(Y)\geq 0, \qquad \forall a\in\mathbb{R},$$

which means

$$(2\operatorname{cov}(X,Y))^2 - 4\operatorname{var}(X)\operatorname{var}(Y)\leq 0.$$

The result follows.

**QED**

---

**Example 7.2.** Binomial Distribution

Let $X\sim \mathrm{BIN}(n,p)$. That is,

$$\mathbb{P}(X=k) = \binom{n}{k}p^k(1-p)^{n-k}, \qquad \forall k\in\{0,\dots,n\},$$

which is the distribution of the number of successes in $n$ independent Bernoulli trials, each with success probability $p$. As a result,

$$X \overset{\mathrm{d}}{=} Y_1+\cdots+Y_n,$$

where

$$Y_1,\dots,Y_n \overset{\mathrm{iid}}{\sim} \mathrm{B}(p).$$

Hence

$$\mathbb{E}(X) = \sum_{i=1}^n \mathbb{E}(Y_i) = np$$

and

$$\operatorname{var}(X) = \sum_{i=1}^n \operatorname{var}(Y_i) = np(1-p).$$

---

# VIII. Convergence of Sequence of Random Variables

## 1. Characteristic Function

Characteristic function serves as a useful tool in determining convergence of random variables.

Def'n 8.1. **Characteristic Function** of a Random Variable
Let $X$ be a random variable. We define the ***characteristic function*** of $X$, denoted as $\varphi_X$, by

$$\varphi_X(t) = \mathbb{E}\left(e^{itX}\right), \qquad \forall t \in \mathbb{R}.$$

Note that

$$\varphi_X(t) = \mathbb{E}(\cos(tX)) + i\,\mathbb{E}(\sin(tX)), \qquad \forall t \in \mathbb{R}.$$

When $X$ admits a pdf $p_X : \mathbb{R} \to \mathbb{R}$, then $\varphi_X$ is the *Fourier transform* of $p_X$.

Unlike the moment generating function $\mathbb{E}\left(e^{tX}\right)$ or the generating function $\mathbb{E}\left(s^X\right)$ of nonnegative integer valued random variables, the characteristic function always exists.

**Proposition 8.1.** Properties of Characteristic Functions

Let $X$ be a random variable.

(a) $\varphi_X(0) = 1$.

(b) $\varphi_X(-t) = \overline{\varphi_X(t)}$ for all $t \in \mathbb{R}$.

(c) $\|\varphi\|_\infty \leq 1$.

(d) $\varphi_{aX+b}(t) = e^{itb}\varphi(at)$ for all $t \in \mathbb{R}$.

---

**Example 8.1.** Characteristic Function of Poisson Random Variable

Let $X \sim \text{POI}(\lambda)$. That is,

$$\mathbb{P}(X = x) = \lambda^k \frac{e^{-\lambda}}{k!}, \qquad \forall k \geq 0.$$

Then

$$\varphi_X(x) = \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k e^{itk}}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\left(\lambda e^{it}\right)^k}{k!} = e^{-\lambda}e^{\lambda e^{it}} = e^{\lambda\left(e^{it}-1\right)}.$$

Now suppose $Y \sim \text{POI}(\eta)$ is independent of $X$. Then

$$\varphi_{X+Y}(t) = \mathbb{E}\left(e^{it(X+Y)}\right) = \mathbb{E}\left(e^{itX}e^{itY}\right) = \mathbb{E}\left(e^{itX}\right)\mathbb{E}\left(e^{itY}\right) = \varphi_X(t)\,\varphi_Y(t), \qquad \forall t \in \mathbb{R},$$

by independence, so that

$$\varphi_{X+Y}(t) = e^{(\lambda+\eta)\left(e^{it}-1\right)}, \qquad \forall t \in \mathbb{R}.$$

We observe that $\varphi_{X+Y}$ is the characteristic function of $\text{POI}(\lambda+\eta)$. We ask:

*does $\varphi_{X+Y}$ being the characteristic function of $\text{POI}(\lambda+\eta)$ imply that $X + Y \sim \text{POI}(\lambda+\eta)$?*

---

More generally, we ask:

*does a characteristic function $\varphi_X$ determine the distribution of $X$?*

The answer is positive, due to the following result.

**Theorem 8.2.** Inversion Formula

Let $X$ be a random variable and let

$$\varphi(t) = \int e^{itx}\,\mathbb{P}(dx), \qquad \forall t \in \mathbb{R}.$$

Then for all $a < b$,

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

**Proof.** For convenience, let

$$I_T = \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t)\,dt, \qquad \forall T \in \mathbb{R}.$$

Then note that

$$\begin{aligned} I_T &= \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \int e^{itx}\,\mathbb{P}(dx)\,dt \\ &= \int \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} e^{itx} dt\,\mathbb{P}(dx) && \text{Fubini} \\ &= \int \int_{-T}^{T} \frac{\sin(t(x-a))}{t} dt - \int_{-T}^{T} \frac{\sin(t(x-b))}{t} dt\,\mathbb{P}(dx). \end{aligned}$$

But for all $\theta > 0$,

$$\int_{-T}^{T} \frac{\sin(\theta t)}{t} dt = \int_{-T}^{T} \frac{\sin(\theta t)}{\theta t} d\theta t = \int_{-T\theta}^{T\theta} \frac{\sin(y)}{y} dy \overset{T\to\infty}{\to} \int_{-\infty}^{\infty} \frac{\sin(y)}{y} dy = \pi.$$

Similarly, for $\theta < 0$,

$$\int_{-T}^{T} \frac{\sin(\theta t)}{t} dt \overset{T\to\infty}{\to} -\pi.$$

For $\theta = 0$,

$$\int_{-T}^{T} \frac{\sin(\theta t)}{t} dt \overset{T\to\infty}{\to} = 0.$$

Therefore,

$$\int_{-T}^{T} \frac{\sin(t(x-a))}{t} dt - \int_{-T}^{T} \frac{\sin(t(x-b))}{t} dt \overset{T\to\infty}{\to} \begin{cases} 2\pi & \text{if } a < x < b \\ \pi & \text{if } x = a \text{ or } x = b \\ 0 & \text{if } x < a \text{ or } x > b \end{cases}.$$

Note that

$$\int_{-T}^{T} \frac{\sin(t(x-a))}{t} dt \le \sup_{c>0} \int_{-c}^{c} \frac{\sin(y)}{y} dy = M < \infty, \qquad \forall T > 0,$$

so that

$$\left| \int_{-T}^{T} \frac{\sin(t(x-a))}{dt} - \int_{-T}^{T} \frac{\sin(t(x-b))}{t} dt \right| \le 2M.$$

Hence by the LDCT,

$$I_T \overset{T\to\infty}{\to} 2\pi\mu((a,b)) + \pi\mu(\{a,b\}),$$

from which the result follows.

**QED**

As a corollary to Theorem 8.2 and Example 8.1, we have that,

$$X \sim \mathrm{POI}(\lambda), Y \sim \mathrm{POI}(\eta) \text{ are independent} \implies X + Y \sim \mathrm{POI}(\lambda + \eta).$$

**Example 8.2.** Characteristic Function of Normal Random Variables

Recall that $X$ is a standard normal random variable if and only if

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}, \qquad \forall x \in \mathbb{R} >$$

The characteristic function of $X$ is given by

$$\begin{aligned} \varphi_X(t) = \mathbb{E}\left(e^{itX}\right) &= \int e^{itx} f(x)\, dx = \int \frac{1}{2\pi} e^{itx} e^{-\frac{x^2}{2}}\, dx \\ &= e^{-\frac{t^2}{2}} \int \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-it)^2}{2}}\, dx = e^{\frac{-t^2}{2}} \int \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\, dy = e^{-\frac{t^2}{2}}, \end{aligned}$$

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

For general $Y \sim \mathcal{N}\left(\mu, \sigma^2\right)$,

$$\varphi_Y(t) = e^{i\mu t - \frac{\sigma^2 t^2}{2}}, \qquad \forall t \in \mathbb{R}.$$

In case $Z \sim \mathcal{N}\left(\mu_2, \sigma_2^2\right)$ is independent of $Y$,

$$\varphi_{Y+Z}(t) = e^{i\left(\mu+\mu_2\right)t - \frac{\left(\sigma^2+\sigma_2^2\right)t}{2}}, \qquad \forall t \in \mathbb{R},$$

so that $Y + Z \sim \mathcal{N}\left(\mu + \mu_2, \sigma^2 + \sigma_2^2\right)$.

---

Def'n 8.2. **Normal** Random Vector

Let $X = (X_1, \ldots, X_n)$ be a random vector. We say $X$ is ***normal*** if any linear combination

$$\sum_{j=1}^{n} a_j X_j$$

is a normal random variable (possibly degenerate).

We define

$$\varphi_X(t) = \mathbb{E}\left(e^{i\sum_{j=1}^{n} t_j X_j}\right), \qquad \forall t \in \mathbb{R}^n.$$

**Theorem 8.3.**

Let $X$ be a random vector. Then

$$X \text{ is normal} \iff \varphi_X(t) = e^{i\langle t,\mu\rangle - \frac{1}{2}\langle t,\Sigma t\rangle}, \qquad \forall t \in \mathbb{R}^n,$$

where $\mu \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n\times n}$ is a symmetric positive semidefinite matrix.

If the implication holds, then $\mu = \left(\mathbb{E}(X_1), \ldots, \mathbb{E}(X_n)\right)$ and $\Sigma$ is the covariance matrix of $X$.

**Proof.** ( $\Longleftarrow$ ) Suppose that

$$\varphi_X(t) = e^{i\langle t,\mu\rangle - \frac{1}{2}\langle t,\Sigma t\rangle}, \qquad \forall t \in \mathbb{R}^n.$$

Then for

$$Y = \sum_{j=1}^{n} a_j X_j = \langle a, X\rangle,$$

the characteristic function $\varphi_Y$ is given by

$$\varphi_Y(t) = \mathbb{E}\left(e^{itY}\right) = \mathbb{E}\left(e^{it\langle a,X\rangle}\right) = \mathbb{E}\left(e^{i\langle ta,X\rangle}\right) = \varphi_X(ta) = e^{i\langle ta,\mu\rangle - \frac{1}{2}\langle ta,\Sigma ta\rangle} = e^{it\langle a,\mu\rangle - \frac{t^2}{2}\langle a,\Sigma a\rangle}, \qquad \forall t \in \mathbb{R},$$

which is the characteristic function of a normal distribution $\mathcal{N}\left(a^T\mu, a^T\Sigma a\right)$.

( $\Longrightarrow$ ) Suppose $X$ is normal. The for any $a \in \mathbb{R}^n$, consider

$$Y = a^T X.$$

Then $Y$ is a normal random variable with

$$\mathbb{E}\left(Y\right) = a^T\mu$$

and

$$\operatorname{var}\left(Y\right) = a^T\Sigma a.$$

This means

$$\varphi_Y\left(t\right) = e^{ita^T\mu - \frac{t^2}{2}a^T\Sigma a}, \qquad \forall t \in \mathbb{R}\,.$$

Hence

$$\varphi_X\left(a\right) = \mathbb{E}\left(e^{ia^TX}\right) = \mathbb{E}\left(e^{iY}\right) = \varphi_Y\left(1\right) = e^{ia^T\mu - \frac{1}{2}a^T\Sigma a}.$$

**QED**

In words, the distribution of a normal random variable is completely determined by its mean vector and covariance matrix.

**Corollary 8.3.1.**

Let $X \sim \mathcal{N}\left(\mu, \Sigma\right)$, then for any $A \in \mathbb{R}^{m\times n}$,

$$AX \sim \mathcal{N}\left(A\mu, A\Sigma A^T\right).$$

**Corollary 8.3.2.**

Let $X \sim \mathcal{N}\left(\mu, \Sigma\right)$. Then for all $i, j \in \{1, \ldots, n\}$,

$$X_i, X_j \text{ are independent} \iff \operatorname{cov}\left(X_i, X_j\right) = 0.$$

## 2. Convergence

Def'n 8.3. **Almost Sure** Convergence

We say a sequence $(X_n)_{n=1}^{\infty}$ of random variables converges (pointwise) ***almost surely*** to a random variable $X$ if

$$\mathbb{P}\left(\left\{\omega \in \Omega : \lim_{n\to\infty} X_n\left(\omega\right) = X_n\left(\omega\right)\right\}\right) = 1.$$

We write $X_n \overset{\text{as}}{\to} X$.

Def'n 8.4. Convergence in $\mathcal{L}^p$

We say a sequence $(X_n)_{n=1}^{\infty}$ of random variables converges in $\mathcal{L}^p$ to a random variable $X$ if $X \in \mathcal{L}^p$ (i.e. $\|X\|_p < \infty$) and

$$\lim_{n\to\infty} \mathbb{E}\left(|X_n - X|^p\right) = 0.$$

That is,

$$\|X_n - X\|_p = \left(\int_\Omega |X_n - X|^p \, d\,\mathbb{P}\right)^{\frac{1}{p}} \to 0.$$

We write $X_n \overset{\mathcal{L}^p}{\to} X$.

Observe that

$$X_n \overset{\mathcal{L}^p}{\to} X \implies \mathbb{E}\left(X_n^p\right) \to \mathbb{E}\left(X^p\right).$$

Def'n 8.5. Convergence in **Probability**

We say a sequence $(X_n)_{n=1}^{\infty}$ of random variables converges in ***probability*** to a random variable $X$ if for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon\}\right) = 0.$$

We write $X_n \xrightarrow{\mathbb{P}} X$.

**Proposition 8.4.**

Let $(X_n)_{n=1}^{\infty}$ be a sequence of random variables. Then

(a) $X_n \xrightarrow{\text{as}} X \implies X_n \xrightarrow{\mathbb{P}} X$; and

(b) for every $p \geq 1$, $X_n \xrightarrow{\mathcal{L}^p} X \implies X_n \xrightarrow{\mathbb{P}} X$.

**Proof.** PMATH 451.

**QED**

**Proposition 8.5.**

Let $p < q$ in $[1, \infty)$. Then

$$X_n \xrightarrow{\mathcal{L}^q} X \implies X_n \xrightarrow{\mathcal{L}^p} X.$$

**Proof.** Let $\varepsilon \in (0,1)$. Then

$$\mathbb{E}\left(|X_n - X|^p\right) = \mathbb{E}\left(|X_n - X|^p \chi_{\{|X_n - X| \geq \varepsilon\}}\right) + \mathbb{E}\left(|X_n - X|^p \chi_{\{|X_n - X| < \varepsilon\}}\right).$$

But when $|X_n - X| \geq \varepsilon$,

$$|X_n - X|^p \leq \varepsilon^{p-q} |X_n - X|^q,$$

so that

$$\mathbb{E}\left(|X_n - X|^p\right) \leq \varepsilon^{p-q}\, \mathbb{E}\left(|X_n - X|^q \chi_{\{|X_n - X| \geq \varepsilon\}}\right) + \varepsilon^p \leq \mathbb{E}\left(|X_n - X|^q\right) + \varepsilon^p.$$

Hence

$$\limsup_{n\to\infty} \mathbb{E}\left(|X_n - X|^p\right) = \varepsilon^{p-q} \limsup_{n\to\infty} \mathbb{E}\left(|X_n - X|^q\right) + \varepsilon^p = \varepsilon^p,$$

so by taking $\varepsilon \to 0$, we see that

$$\limsup_{n\to\infty} \mathbb{E}\left(|X_n - X|^p\right) = 0.$$

**QED**

In general, convergence in probability does not imply convergence in $L^p$ or almost sure convergence.

**Example 8.3.**

Consider $((0,1), \mathcal{B}((0,1)), m)$, where $m$ is the Lebesgue measure. For all $n \in \mathbb{N}$, consider $k_n \in \mathbb{N}$ such that $1 + \cdots + k_n < n \leq 1 + \cdots + (k_n + 1)$. Define

$$\begin{aligned} X_n : (0,1) &\to \mathbb{R} \\ x &\mapsto \begin{cases} 1 & \text{if } n - k_n - 1 < x < n - k_n \\ 0 & \text{otherwise} \end{cases}. \end{aligned}$$

Then observe that $X_n \to 0$ in probability and in $L^p$ (for all $p \geq 1$), but $X_n$ does not converge almost surely to 0.

**Example 8.4.**

Again consider $((0,1), \mathcal{B}((0,1)), m)$ with

$$X_n : (0,1) \to \mathbb{R}$$
$$x \mapsto \begin{cases} n & \text{if } x < \frac{1}{n} \\ 0 & \text{otherwise} \end{cases}.$$

Then observe that $X_n \to 0$ in probability and almost surely, but $X_n$ does not converge to 0 in $L^p$.

---

**Proposition 8.6.**

Let $(X_n)_{n=1}^{\infty}$ be a sequence of random variables such that $X_n \to X$ in probability. Then there exists a subsequence $(X_{n_k})_{k=1}^{\infty}$ such that $X_{n_k} \to X$ almost surely.

**Proof.** PMATH 451.

**QED**

Def'n 8.6. **Uniformly Integrable** Collection of Random Variables

We say a collection $\mathcal{C}$ of random variables is ***uniformly integrable*** if

$$\lim_{k\to\infty} \sup_{f\in\mathcal{C}} \int_{|X|>k} |X| \, d\mathbb{P} = \lim_{k\to\infty} \sup_{f\in\mathcal{C}} \mathbb{E}\left(|X| \chi_{|X|>k}\right) = 0$$

**Proposition 8.7.**

Let $(X_n)_{n=1}^{\infty}$ be a sequence of random variables such that $X_n \to X$ in probability. If $(X_n)_{n=1}^{\infty}$ is uniformly integrable in addition, then $X_n \to X$ in $L^1$.

**Proof.** Without loss of generality, assume $X = 0$.

Let $\varepsilon > 0$ be fixed. Since $(X_n)_{n=1}^{\infty}$ is uniformly integrable, there is $t_\varepsilon > 0$ such that

$$\mathbb{E}\left(|X_n| \chi_{|X_n|>t_\varepsilon}\right) < \varepsilon, \qquad \forall n \in \mathbb{N}.$$

Since $X_n \to 0$ in probability, there is $N_\varepsilon$ such that for all $n \geq N_\varepsilon$,

$$\mathbb{P}(|X_n| > \varepsilon) < \frac{\varepsilon}{t_\varepsilon}.$$

Hence for $n \geq N_\varepsilon$,

$$\mathbb{E}(|X_n|) = \int_{\{|X_n|\leq\varepsilon\}} |X_n| \, d\mathbb{P} + \int_{\{\varepsilon<|X_n|\leq t_\varepsilon\}} |X_n| \, d\mathbb{P} + \int_{\{|X_n|>t_\varepsilon\}} |X_n| \, d\mathbb{P} \leq \int_{\{|X_n|\leq\varepsilon\}} \varepsilon d\mathbb{P} + t_\varepsilon \mathbb{P}(|X_n| > \varepsilon) + \varepsilon \leq 3\varepsilon.$$

Since $\varepsilon > 0$ was chosen arbitrarily, we conclude that

$$\mathbb{E}(|X_n|) \to 0,$$

as required.

**QED**

Def'n 8.7. Convergence in **Distribution**

We say a sequence $(X_n)_{n=1}^{\infty}$ of random variables with pdfs $F_1, F_2, \ldots : \mathbb{R} \to [0,1]$ converges in ***distribution*** to a random variable $X$ with pdf if $F_n \to F$ pointwise except possibly at points where $F$ is discontinuous.

**Example 8.5.**

Consider independent trials until the first success, with success probability $p$. The the number of trials $X_p$ until a success has geometric distribution with parameter $p$:

$$\mathbb{P}\left(X_p = n\right) = p(1-p)^{n-1}, \qquad \forall n \in \mathbb{N}.$$

This means

$$\mathbb{P}\left(X_p > n\right) = (1-p)^n.$$

As $p \to 0$,

$$\lim_{p\to 0} \mathbb{P}\left(pX_p > x\right) = \lim_{p\to 0} \mathbb{P}\left(X_p > \frac{x}{p}\right) = \lim_{p\to 0} (1-p)^{\frac{x}{p}} = e^{-x}, \qquad \forall x \geq 0.$$

This means

$$\mathbb{P}\left(pX_p \leq x\right) \to 1 - e^{-x}, \qquad \forall x \geq 0,$$

which is the pdf of exponential distribution.

---

**Proposition 8.8.**
Let $(X_n)_{n=1}^{\infty}$ be a sequence of random variables. If $X_n \to X$ in probability, then $X_n \to X$ in distribution.

---

**Proposition 8.9.**
Let $(X_n)_{n=1}^{\infty}$ be a sequence of random variables such that $X_n \to c$ in distribution for some constant $c \in \mathbb{R}$. Then $X_n \to c$ in probability.

**Proof.** For any $\varepsilon > 0$, observe that

$$\mathbb{P}\left(|X_n - c| \leq \varepsilon\right) = \mathbb{P}\left(c - \varepsilon \leq X_n \leq c + \varepsilon\right) \geq \mathbb{P}\left(c - \varepsilon < X_n \leq c + \varepsilon\right) = F_{X_n}\left(c + \varepsilon\right) - F_{X_n}\left(c - \varepsilon\right), \qquad \forall n \in \mathbb{N}.$$

Since $F_c = \chi_{[c,\infty)}$, so that $F_n\left(c+\varepsilon\right) \to F_c\left(c+\varepsilon\right) = 1$ and $F_n\left(c-\varepsilon\right) = F_c\left(c-\varepsilon\right) = 0$. Thus

$$\liminf_{n\to\infty} \mathbb{P}\left(|X_n - c| \leq \varepsilon\right) \geq 1,$$

which imply that

$$\lim_{n\to\infty} \mathbb{P}\left(|X_n - c| \leq \varepsilon\right) = 1.$$

**QED**

---

**Theorem 8.10.** Skorokhod's Theorem
Let $(X_n)_{n=1}^{\infty}$ be a sequence of random variables such that $X_n \to X$ in distribution. Then there is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a sequence of random variables $(Y_n)_{n=1}^{\infty}$ such that $Y_n \to Y$ for some random variable $Y$ on $(\Omega, \mathcal{F}, \mathbb{P})$, $X_n = Y_n, X = Y$ in distribution.

---

Def'n 8.8. **$\mu$-continuity Set**
Let $(X, \mathcal{A}, \mu)$ be a measure space, where $X$ is a topological space. We say $A \in \mathcal{A}$ is a ***$\mu$-continuity set*** if

$$\mu\left(\partial A\right) = 0.$$

**Theorem 8.11.** Portmanteau Theorem
Let $(X_n)_{n=1}^{\infty}$ be a sequence of random variables and let $X$ be a random variable. The following are equivalent.

(a) $X_n \to X$ in distribution.

(b) $\mathbb{E}\left(f(X_n)\right) \to \mathbb{E}\left(f(X)\right)$ for all bounded continuous function $f: \mathbb{R} \to \mathbb{R}$.

(c) $\mu_{X_n}(A) \to \mu_X(A)$ for every $\mu$-continuity set $A \in \mathrm{Bor}(\mathbb{R})$.

---

**Theorem 8.12.** Helly's Selection Theorem
Let $(F_n)_{n=1}^{\infty}$ be a sequence of distribution functions. Then there is a subsequence $(F_{n_k})_{k=1}^{\infty}$ and a right-continuous nondecreasing function $F$ such that

$$\lim_{k\to\infty} F_{n_k}(y) = F(y)$$

for all $y$ at which $F$ is continuous.

**Proof.** Assume we count the rational numbers as $\{q_j\}_{j=1}^{\infty}$. Since $(F_n(q_1))_{n=1}^{\infty}$ is bounded, there is a subsequence $\left(F_{m_1(k)}\right)_{k=1}^{\infty}$ such that

$$F_{m_1(k)}(q_1) \to G(q_1) \in \mathbb{R}.$$

Take a subsequence $\left(F_{m_2(k)}\right)_{k=1}^{\infty}$ of $\left(F_{m_1(k)}\right)_{k=1}^{\infty}$ so that

$$F_{m_2(k)}(q_2) \to G(q_2) \in \mathbb{R}.$$

Continue this process to obtain subsequences. Then we can use a diagonal argument as follows. Note that we have

$$\begin{array}{cccc} F_{m_1(1)} & F_{m_1(2)} & F_{m_1(3)} & \cdots \\ F_{m_2(1)} & F_{m_2(2)} & F_{m_2(3)} & \cdots \\ F_{m_3(1)} & F_{m_3(2)} & F_{m_3(3)} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{array}$$

Take the *diagonal* sequence $\left(F_{m_k(k)}\right)_{k=1}^{\infty}$; then observe that

$$F_{m_k(k)}(q) \to G(q), \qquad \forall q \in \mathbb{Q}.$$

Then $G$ is non-decreasing but it may be the case that $G$ is not right-continuous. To resolve this, take

$$F(x) = \inf_{q \in \mathbb{Q}, q > x} G(q), \qquad \forall x \in \mathbb{R}.$$

Then $F$ is non-decreasing and right-continuous with $F_{m_k(k)}(y) \to F(y)$ for all $y$ at which $F$ is continuous.

**QED**

Note that the limit function $F$ we obtain from Helly's selection theorem may not be a distribution function.

Def'n 8.9. **Tight** Family of Probability Measures on $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}))$

Let $\{\mu_n\}_{n=1}^{\infty}$ be a family of probability measures on $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}))$. We say $\{\mu_n\}_{n=1}^{\infty}$ is ***tight*** if for every $\varepsilon > 0$, there is $M > 0$ such that

$$\liminf_{n\to\infty} \mu_n([-M, M]) \geq 1 - \varepsilon.$$

Note that

$$\liminf_{n\to\infty} \mu_n([-M, M]) \geq 1 - \varepsilon \iff \limsup_{n\to\infty} 1 - F_n(M) + F_n(-M) < \varepsilon.$$

**Theorem 8.13.**

Let $\{\mu_n\}_{n=1}^{\infty}$ be a family of probability measures on $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}))$. Then $\{\mu_n\}_{n=1}^{\infty}$ is tight if and only if it admits a subsequence that converges weakly to a probability measure.

**Proof.** ($\Longrightarrow$) By Helly's selection theorem, there is a subsequence $(F_{n_k})_{k=1}^{\infty}$ such that $F_{n_k} \to F$ at points where $F$ is continuous. It remains to show that $\lim_{y\to-\infty} F(y) = 0$ and $\lim_{y\to\infty} F(y) = 1$.

Since $F_{n_k} \to F$ ae,

$$0 \leq \lim_{y\to-\infty} F(y) \leq \lim_{y\to\infty} F(y) \leq 1.$$

Hence it suffices to show that

$$\lim_{y\to\infty} F(y) - F(-y) = 1.$$

For $\varepsilon > 0$, define $M_\varepsilon$ such that

$$\limsup_{n\to\infty} 1 - F_n(M_\varepsilon) + F_n(-M_\varepsilon) \leq \varepsilon.$$

Take $r < -M_\varepsilon, s > M_\varepsilon$ to be continuity points of $F$. Then

$$1 - F(s) + F(r) = \lim_{k\to\infty} 1 - F_{n_k}(s) + F_{n_k}(r) \le \limsup_{n\to\infty} 1 - F_n(M_\varepsilon) + F_n(-M_\varepsilon) \le \varepsilon.$$

Hence

$$\lim_{y\to\infty} F(y) - F(-y) \ge F(s) - F(r) \ge 1 - \varepsilon$$

holds for any $\varepsilon > 0$, so that $\lim_{y\to\infty} F(y) - F(-y) = 1$, as needed.

( $\Longleftarrow$ ) Suppose $\{\mu_n\}_{n=1}^\infty$ is not tight. Then there is $\varepsilon > 0$ and a subsequence $(F_{n_k})_{k=1}^\infty$ such that

$$1 - F_{n_k}(M) + F_{n_k}(-M) \ge \varepsilon, \qquad \forall M > 0.$$

By Helly's selection, there exist a subsequence $(F_{n_k})_{k=1}^\infty$ such that $F_{n_k}(x) \to F(x)$ for all continuity point $x$ of $F$. Let $r < 0 < s$ be continuity points of $F$. Then

$$1 - F(s) + F(r) = \lim_{k\to\infty} 1 - F_{n_k}(s) + F_{n_k}(r) \ge \liminf_{k\to\infty} 1 - F_{n_k}(s) + F_{n_k}(r) \ge \varepsilon.$$

Taking $r \to -\infty, s \to \infty$, we obtain

$$F(\infty) - F(-\infty) \le 1 - \varepsilon < 1.$$

Thus $F$ is not a distribution function.

**QED**

**Proposition 8.14.**
Let $\varphi$ be the characteristic function of a probability measure $\mu$ on $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}))$, then $\varphi$ is continuous on $\mathbb{R}$.

**Proposition 8.15.**
If $(\mu_n)_{n=1}^\infty$ is a tight sequence of probability measure such that every weakly convergent subsequence converges to a same limit $\mu$, then $\mu_n \to \mu$ weakly.

**Proof.** Suppose $\mu_n \not\to \mu$ weakly. Then there is a point $x$ at which the distribution $F$ of $\mu$ is continuous and $F_n(x) \not\to F(x)$, where each $F_n$ is the distribution of $\mu_n$. Hence there is $\varepsilon > 0$ such that

$$|F_n(x) - F(x)| \ge \varepsilon$$

for infinitely many $n$. Hence take a subsequence $\left(\mu_{n_k}\right)_{k=1}^\infty$ such that

$$|F_{n_k}(x) - F(x)| \ge \varepsilon$$

for all $k \in \mathbb{N}$. Since $(\mu_n)_{n=1}^\infty$ is tight, so is $\left(\mu_{n_k}\right)_{k=1}^\infty$. By tightness, there is another subsequence $\left(\mu_{n_{k_j}}\right)_{j=1}^\infty$ that converges weakly. But then

$$\left|F_{n_{k_j}}(x) - F(x)\right| \ge \varepsilon, \qquad \forall j \in \mathbb{N},$$

which contradicts assumption that every weakly convergent subsequence of $(\mu_n)_{n=1}^\infty$ converges to $\mu$.

**QED**

**Theorem 8.16.** Continuity Theorem
Let $\mu_1, \mu_2, \ldots, \mu$ be probability measures on $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}))$ with characteristic functions $\varphi_1, \varphi_2, \ldots, \varphi$, respectively. Then

$$\mu_n \to \mu \text{ weakly} \iff \varphi_n \to \varphi \text{ pointwise}.$$

**Proof.** ( $\Longrightarrow$ ) Suppose $\mu_n \to \mu$ weakly. For a fixed $t \in \mathbb{R}$, note that $e^{itx}$ is a continuous bounded function of $x$, so by the Portmanteau theorem,

$$\varphi_n(t) = \int e^{itx} d\mu_n(x) \to \int e^{itx} d\mu(x) = \varphi(t).$$

But this precisely means $\varphi_n \to \varphi$ pointwise, since the choice of $t$ was arbitrary.

( $\Longleftarrow$ ) We consider the following claim.

Claim 1. *$(\mu_n)_{n=1}^{\infty}$ is tight.*

Note that, given $u > 0$,

$$\begin{aligned}\frac{1}{u}\int_{-u}^{u} 1-\varphi_n(t)\,dt &= \int \frac{1}{u}\int_{-u}^{u} 1-e^{itx} dt d\mu_n(x) = 2\int 1-\frac{\sin(ux)}{ux} d\mu_n(x) \\ &\geq 2\int_{\left(-\infty,-\frac{2}{u}\right]\cup\left[\frac{2}{u},\infty\right)} 1-\frac{1}{|ux|} d\mu_n(x) \geq \mu_n\left(\left(-\infty,-\frac{2}{u}\right]\cup\left[\frac{2}{u},\infty\right)\right)\end{aligned}$$

Since $\varphi$ is continuous and $\varphi(0) = 1$,

$$\lim_{u\to 0}\frac{1}{u}\int_{-u}^{u}(1-\varphi(t))\,dt = 0.$$

Hence for any $\varepsilon > 0$, there is $u > 0$ such that

$$\frac{1}{u}\int_{-u}^{u}(1-\varphi(t))\,dt < \frac{\varepsilon}{2}.$$

Since $\varphi_n \to \varphi$ pointwisely and $1-\varphi_n \leq 1$, by the Lebesgue dominated convergence theorem,

$$\frac{1}{u}\int_{-u}^{u}\left(1-\varphi_n(t)\right)dt \to \frac{1}{u}\int_{-u}^{u}(1-\varphi(t))\,dt.$$

So there is $N \in \mathbb{N}$ such that for all $n \geq N$,

$$\frac{1}{u}\int_{-u}^{u} 1-\varphi_n(t)\,dt < \varepsilon.$$

Hence

$$\mu_n\left(\left(-\infty,-\frac{2}{u}\right]\cup\left[\frac{2}{u},\infty\right)\right) < \varepsilon$$

for all $n \geq N$, which implies

$$\limsup_{n\to\infty}\mu_n\left(\left(-\infty,-\frac{2}{u}\right]\cup\left[\frac{2}{u},\infty\right)\right) < \varepsilon.$$

Since the choice of $\varepsilon$ was arbitrary, it follows $(\mu_n)_{n=1}^{\infty}$ is tight.

*(End of Claim 1)*

Assume a subsequence $\left(\mu_{n_k}\right)_{k=1}^{\infty}$ converges weakly to $\mu'$. By the forward direction, we konw that $\varphi_{n_k} \to \varphi'$ pointwise, where $\varphi'$ is the characteristic function of $\mu'$. But we know that $\varphi_n \to \varphi$ pointwise, so that $\varphi = \varphi'$. Since the characteristic function completely determines the distribution, we conclude that $\mu = \mu'$. It follows that every weakly convergent subsequence converges to $\mu$, so by Proposition 8.12, $\mu_n \to \mu$ weakly, as needed.

**QED**

# IX. Limit Theorems

## 1. Weak Law of Large Numbers

**Theorem 9.1.** WLLN

Let $(X_n)_{n=1}^{\infty}$ be a sequence of uncorrelated random variables with $\operatorname{var}(X_n) \leq c$ for some $c \geq 0$ and the same expectation. Let

$$S_n = \sum_{k=1}^{n} X_k$$

for all $n \in \mathbb{N}$. Then $\frac{S_n}{n} \to \mu$ in $L^2$ and in probability.

**Proof.** Note that $\mathbb{E}\left(\frac{S_n}{n}\right) = \mu$ for all $n \in \mathbb{N}$.

Also,

$$\operatorname{var}\left(\frac{S_n}{n}\right) = \mathbb{E}\left(\left(\frac{S_n}{n} - \mu\right)^2\right) = \frac{1}{n^2}\sum_{k=1}^{n} \operatorname{var}(X_k) \leq \frac{c}{n} \to 0.$$

This means $\frac{S_n}{n} \to \mu$ in $L^2$ and in probability.

**QED**

**Theorem 9.2.** WLLN for Triangular Arrays

Consider a trangular array of random variables,

$$\begin{array}{cccc} X_{1,1} & & & \\ X_{2,1} & X_{2,2} & & \\ X_{3,1} & X_{3,2} & X_{3,3} & \\ \vdots & \vdots & \vdots & \ddots \end{array},$$

and let

$$S_n = \sum_{j=1}^{n} X_{n,j}.$$

If there is a sequence $(b_n)_{n=1}^{\infty} \in \mathbb{R}^{\mathbb{N}}$ such that

$$\frac{\operatorname{var}(S_n)}{b_n^2} \to 0,$$

then

$$\frac{S_n - \mathbb{E}(S_n)}{b_n} \to 0$$

in probability.

**Proof.** Note

$$\mathbb{E}\left(\left(\frac{S_n - \mathbb{E}(S_n)}{b_n}\right)^2\right) = \frac{1}{b_n^2}\,\mathbb{E}\left((S_n - \mathbb{E}(S_n))^2\right) = \frac{\operatorname{var}(S_n)}{b_n^2} \to 0.$$

Hence $\frac{S_n - \mathbb{E}(S_n)}{b_n} \to 0$ in probability.

**QED**

If we further assume that $X_{n,j}$'s have identical distribution with mean and variance $\mu, \sigma^2$ and the random variables in the same row are independent, then

$$\frac{S_n}{n} \to \mu$$

in probability.

For sequences of random variables without finite second moment, we consider the following.

**Theorem 9.3.**
Let $\left(X_{n,k}\right)_{1\le k\le n}$ be a (triangular) sequence of random variables, such that $X_{n,1},\dots,X_{n,n}$ are independent for all $n\in\mathbb{N}$. Let $(b_n)_{n=1}^{\infty}\in(0,\infty)^{\mathbb{N}}$ such that $b_n\to\infty$ and let

$$\overline{X_{n,k}}=X_{n,k}\chi_{\left|X_{n,k}\right|\le b_n},\qquad\qquad \forall n\in\mathbb{N},k\le n.$$

Suppose $\sum_{k=1}^{n}\mathbb{P}\left(\left|X_{n,k}\right|>b_n\right)\to 0$ and that $b_n^{-2}\sum_{k=1}^{n}\mathbb{E}\left(\overline{X_{n,k}}^2\right)\to 0$, then

$$\frac{S_n-a_n}{b_n}\to 0$$

in probability, where

$$S_n=\sum_{k=1}^{n}X_{n,k}$$

and

$$a_n=\sum_{k=1}^{n}\mathbb{E}\left(\overline{X_{n,k}}\right).$$

---

**Theorem 9.4.**
Let $X_1,X_2,\dots$ be iid with $\lim_{x\to\infty}x\,\mathbb{P}\left(|X_1|>x\right)=0$. Let

$$S_n=\sum_{k=1}^{n}X_k$$

and

$$\mu_n=\mathbb{E}\left(X_1\chi_{|X_1|\le n}\right)$$

for all $n\in\mathbb{N}$. Then

$$\frac{S_n}{n}-\mu_n\to 0\text{ in probability.}$$

---

**Lemma 9.5.**
Let $Y$ be a nonnegative random variable. Then

$$\mathbb{E}\left(Y^2\right)=\int_0^{\infty}2y\,\mathbb{P}\left(Y>y\right)dy.$$

**Proof.** Recall that

$$\mathbb{P}\left(Y>y\right)=\mathbb{E}\left(\chi_{Y>y}\right)=\int_0^{\infty}\mathbb{P}\left(Y>y\right)dy.$$

Note

$$\int_0^{\infty}2y\,\mathbb{P}\left(Y>y\right)dy=\int_0^{\infty}\mathbb{E}\left(2y\chi_{Y>y}\right)dy=\mathbb{E}\left(\int_0^{\infty}2y\chi_{Y>y}dy\right)=\mathbb{E}\left(\int_0^{Y}2ydy\right)=\mathbb{E}\left(Y^2\right).$$

**QED**

**Theorem 9.6.**
Let $X_1,X_2,\dots$ be iid and $\mathrm{L}^1$ and let $S_n=\sum_{k=1}^{n}X_k$ for all $n\in\mathbb{N}$. Let $\mu=\mathbb{E}\left(X_1\right)$. Then $\frac{S_n}{n}\to\mu$ in $\mathbb{P}$.

---

**Theorem 9.7.** Strong Law of Large Numbers

Let $X_1, X_2, \ldots$ be identically distributed $L^1$ random variables that are pairwisely independent. Let $\mu = \mathbb{E}(X_1)$ and let $S_n = \sum_{k=1}^{n} X_k$ for all $n \in \mathbb{N}$. Then

$$\frac{S_n}{n} \to \mu$$

almost surely.

---

Suppose $X_1, X_2, \ldots$ are iid, $\mathbb{E}\left(X_1^+\right) = \infty$, and $\mathbb{E}\left(X_1^-\right) < \infty$. Then

$$\frac{S_n}{n} \to \infty.$$

The idea for the proof is that we can use trunctation to show that

$$\liminf_{n\to\infty} \frac{S_n}{n} > x$$

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

**Example 9.1.** Renewal Theory

Consider iid interarrival times $X_1, X_2, \ldots$, such as arrival times of customers, service times of servers, and lifespan of lightbulbs, are positive random variables. For $n \in \mathbb{N}$, consider

$$T(n) = \sum_{j=1}^{n} X_j,$$

the time of the $n$th occurance of the event and for $t \geq 0$, let

$$N(t) = \sup\{n \in \mathbb{N} : T(n) \leq t\},$$

the number of occurance by time $t$. Then the SLLN implies the following.

If $\mathbb{E}(X_1) = \mu \leq \infty$, then

$$\frac{N_t}{t} \to \frac{1}{\mu} \text{ almost surely}$$

as $t \to \infty$.

**Proof.** By the SLLN,

$$\frac{T(n)}{n} \to \mu \text{ almost surely}$$

and note that

$$T(N_t) \leq t < T(N_t + 1), \qquad \forall t \geq 0,$$

so that

$$\frac{T(N_t)}{N_t} \leq \frac{t}{N_t} < \frac{T(N_t+1)}{N_t} = \frac{T(N_t+1)}{N_t+1}\frac{N_t+1}{N_t}.$$

As $t \to \infty$, $N_t \to \infty$ almost surely, so that $\frac{N_t+1}{N_t} \to 1$ almost surely. This means

$$\frac{t}{N_t} \to \mu \text{ almost surely}$$

as $t \to \infty$.

**QED**

**Example 9.2.** Empirical Distribution Function

Let $X_1, X_2, \ldots$ be iid samples from (unknown) distribution function $F$ and let $F_n : \mathbb{R} \to [0,1]$ be

$$F_n(x) = \frac{1}{n} \sum_{m=1}^{n} \chi_{X_m \le x}$$

be the *empirical distribution function*. Then $F_n \to F$ almost surely by SLLN.

As it turns out, the Glivenko-Contelli theorem shows a stronger result:

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \to 0 \text{ almost surely.}$$

That is, we have an almost sure *uniform* convergence.

---

## 3. Central Limit Theorem

**Theorem 9.8.** Central Limit Theorem

Let $X_1, X_2, \ldots$ be iid with $\mathbb{E}(X_1) = \mu$, $\operatorname{var}(X_1) = \sigma^2 \in (0, \infty)$ and let $S_n = \sum_{j=1}^{n} X_j$ for all $n \in \mathbb{N}$. Then

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \to \mathcal{N}(0,1) \text{ in distribution.}$$

---

**Example 9.3.** Normal Approximation of Binomial Distribution

Let $X_1, X_2, \ldots \overset{\text{iid}}{\sim} \mathrm{B}(p)$. Then

$$S_n = \sum_{j=1}^{n} X_j \sim \mathrm{BIN}(n, p), \qquad \forall n \in \mathbb{N}$$

and $\mathbb{E}(X_1) = p$, $\operatorname{var}(X_1) = p(1-p)$.

By the CLT,

$$\frac{S_n - np}{\sqrt{p(1-p)}\sqrt{n}} \to \mathcal{N}(0,1) \text{ in distribution,}$$

which means

$$\frac{S_n - np}{\sqrt{n}} \to \mathcal{N}(0, p(1-p)) \text{ in distribution.}$$

It follows that

$$S_n - np \approx \mathcal{N}(0, np(1-p)),$$

so that

$$S_n \approx \mathcal{N}(np, np(1-p)).$$

In words, we can approximate $\mathrm{BIN}(n,p)$ with $\mathcal{N}(np, np(1-p))$ when $n$ is large.

For instance, when $p = \frac{1}{2}$,

$$\mathbb{P}\left(\frac{S_n - np}{\sqrt{p(1-p)}\sqrt{n}} \in [a,b]\right) \approx \Phi(b) - \Phi(a),$$

where $\Phi$ is the distribution of $\mathcal{N}(0,1)$, which means

$$\mathbb{P}(S_n \in [c,d]) \approx \Phi\left(\frac{d - \frac{n}{2}}{\frac{1}{2}\sqrt{n}}\right) - \Phi\left(\frac{c - \frac{n}{2}}{\frac{1}{2}\sqrt{n}}\right).$$

**Theorem 9.9.** LF CLT

For $n \in \mathbb{N}$, let $X_{n,1}, \ldots, X_{n,n}$ be independent with $\mathbb{E}\left(X_{n,m}\right) = 0$. If

$$\sum_{m=1}^{n} \mathbb{E}\left(X_{n,m}^2\right) \to \sigma^2 > 0$$

and

$$\lim_{n\to\infty} \sum_{m=1}^{n} \mathbb{E}\left(\left|X_{n,m}\right|^2 : \left|X_{n,m}\right| > \varepsilon\right) = 0,$$

then

$$S_n = \sum_{m=1}^{n} X_{n,m} \to \sigma \mathcal{N}\left(0,1\right) \text{ in distribution.}$$

# X. Conditional Expectation

## 1. Conditional Expectation

Def'n 10.1. **Conditional Expectation** of a Random Variable Given a Sub-$\sigma$-field
Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X$ be a random variable on it, with $\mathbb{E}(|X|) < \infty$. Let $\mathcal{G}$ be a sub-$\sigma$-field of $\mathcal{F}$. Then the ***conditional expectation*** of $X$ given $\mathcal{G}$, denoted as $\mathbb{E}(X|\,\mathcal{G})$, is a random variable such that

(a) $\mathbb{E}(X|\,\mathcal{G})$ is $\mathcal{G}$-measurable; and

(b) for any $A \in \mathcal{G}$,

$$\int_A X d\mathbb{P} = \int_A \mathbb{E}(X|\,\mathcal{G})\, d\mathbb{P}\,.$$

To show conditional expectation exists, we recall the Radon-Nikodym theorem from measure theory.

**Theorem 10.1.** Radon-Nikodym Theorem
Let $(X, \mathcal{A})$ be a measurable space and let $\mu, \nu$ be $\sigma$-finite measures on it, such that $\nu \ll \mu$. Then there exists $f \in \mathrm{L}^1(X, \mathcal{A}, \mu)$ such that

$$\int_A f d\mu = \nu(A)\,, \qquad \forall A \in \mathcal{A}\,.$$

---

We often denote $f$ by $\frac{d\nu}{d\mu}$ and is called the Radon-Nikodym derivative or density.

**Proposition 10.2.**
Conditional expectaion exists and is unique almost surely.

**Proof of Existence.** It suffices to consider the case where $X \geq 0$.
Let $\mu = \mathbb{P}|_{\mathcal{G}}$ and let

$$\nu(A) = \int_A X d\mathbb{P}, \qquad \forall A \in \mathcal{G}\,.$$

Then $\mu(A) = 0 \implies \mathbb{P}(A) = 0 \implies \nu(A) = 0$, so that $\nu \ll \mu$. So by the Radon-Nikodym theorem, there exists $Y = \frac{d\nu}{d\mu}$ such that

$$\int_A X d\mathbb{P} = \nu(A) = \int_A Y d\mathbb{P}, \qquad \forall A \in \mathcal{G}\,.$$

**Proof of Uniqueness.** Assume $Y, Y'$ are conditional expectations but $\mathbb{P}(Y \neq Y') > 0$. Without loss of generality, assume $A = \{Y > Y'\}$ has positive probability. Since $Y, Y'$ are $\mathcal{G}$-measurable, $Y, Y' \in \mathcal{G}$. But

$$\int_A Y d\mathbb{P} = \int_A X d\mathbb{P} = \int_A Y' d\mathbb{P},$$

which is a contradiction.

**QED**

**Proposition 10.3.** Properties of Conditional Expectation
Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-$\sigma$-field of $\mathcal{F}$.

(a) For random variables $X, Y$ on $(\Omega, \mathcal{F}, \mathbb{P})$ and $a \in \mathbb{R}$,

$$\mathbb{E}(aX + Y|\,\mathcal{G}) = a\,\mathbb{E}(X|\,\mathcal{G}) + \mathbb{E}(Y|\,\mathcal{G})\,.$$

(b) If $X, Y$ are random variables with $X \leq Y$ almost surely, then

$$\mathbb{E}(X|\mathcal{G}) \leq \mathbb{E}(Y|\mathcal{G}).$$

(c) If $(X_n)_{n=1}^{\infty}$ is an increasing sequence of nonnegative random variables converging almost surely to $X$ with $\mathbb{E}(X) < \infty$, then

$$\mathbb{E}(X_n|\mathcal{G}) \uparrow \mathbb{E}(X|\mathcal{G}).$$

---

In case $\mathcal{G} = \mathcal{F}$, we have

$$\mathbb{E}(X|\mathcal{F}) = X \text{ almost surely.}$$

**Theorem 10.4.**

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{F}_1 \subseteq \mathcal{F}_2$ be sub-$\sigma$-fields of $\mathcal{F}$. Let $X$ be a random variable.

(a) $\mathbb{E}(\mathbb{E}(X|\mathcal{F}_1)|\mathcal{F}_2) = \mathbb{E}(X|\mathcal{F}_1)$.

(b) $\mathbb{E}(\mathbb{E}(X|\mathcal{F}_2)|\mathcal{F}_1) = \mathbb{E}(X|\mathcal{F}_1)$.

**Proof.** Proof of (a) is trivial.

For $A \in \mathcal{F}_1$, $A \in \mathcal{F}_2$ as well, so that

$$\int_A \mathbb{E}(X|\mathcal{F}_1)\,d\mathbb{P} = \int_A X d\mathbb{P} = \int_A \mathbb{E}(X|\mathcal{F}_2)\,d\mathbb{P}.$$

Hence

$$\mathbb{E}(\mathbb{E}(X|\mathcal{F}_2)|\mathcal{F}_1) = \mathbb{E}(X|\mathcal{F}_1).$$

QED

---

In words, note that the *fineness* of $\mathcal{F}_1, \mathcal{F}_2$ can be thought as the *resolution* through which we observe the space $(X, \mathcal{F}, \mathbb{P})$. Hence, no matter the order which we observe the space, we always end up with the worst resolution $\mathcal{F}_1$.

In particular,

$$\mathbb{E}(\mathbb{E}(X|\mathcal{G})|\{\emptyset, \Omega\}) = \mathbb{E}(X|\{\emptyset, \Omega\}) = \mathbb{E}(X)$$

for any sub-$\sigma$-field $\mathcal{G} \subseteq \mathcal{F}$, which is called the *law of iterated expectation*.

**Theorem 10.5.**

Let $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-$\sigma$-field. If $X, Y$ are random variables such that $X, XY$ have finite expectation, then

$$\mathbb{E}(XY|\mathcal{G}) = X\,\mathbb{E}(Y|\mathcal{G}).$$

**Proof.** It suffices to check that $X\,\mathbb{E}(Y|\mathcal{G})$ satisfies the definition of conditional expectation for $\mathbb{E}(Y|\mathcal{G})$. By assumption, it is clear that $X\,\mathbb{E}(Y|\mathcal{G})$ is $\mathcal{G}$-measurable.

For $A, B \in \mathcal{G}$,

$$\int_A \chi_B\,\mathbb{E}(Y|\mathcal{G})\,d\mathbb{P} = \int_{A\cap B} \mathbb{E}(Y|\mathcal{G})\,d\mathbb{P} = \int_{A\cap B} Y d\mathbb{P} = \int_A \chi_B Y d\mathbb{P}.$$

It follows by induction that the result holds when $X$ is a simple function. Then by using simple approximation, we can prove the result for nonnegative random variables, which can be easily extend for general random variables.

QED

---

**Theorem 10.6.**

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a measure space and let $X, Y$ be independent random variables with $\mathbb{E}(|Y|) < \infty$. Then

$$\mathbb{E}(Y|X) = \mathbb{E}(Y)$$

almost surely, where $\mathbb{E}(Y|X) = \mathbb{E}(Y|\sigma(X))$.

**Proof.** Recall that

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

This means, for $A \in \sigma(X)$,

$$\int_A Y d\mathbb{P} = \int Y\chi_A d\mathbb{P} = \mathbb{E}\left(Y\chi_A\right) = \mathbb{E}(Y)\,\mathbb{E}\left(\chi_A\right),$$

since $X, Y$ are independent. But

$$\mathbb{E}(Y)\,\mathbb{E}\left(\chi_A\right) = \int_A \mathbb{E}(Y)\, d\mathbb{P},$$

so that

$$\int_A \mathbb{E}(Y|X)\, d\mathbb{P} = \int_A Y d\mathbb{P} = \int_A \mathbb{E}(Y)\, d\mathbb{P}, \qquad \forall A \in \sigma(X),$$

which means $\mathbb{E}(Y|X) = \mathbb{E}(Y)$ almost surely.

**QED**

**Theorem 10.7.** Jensen's Inequality - Conditional Version

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-$\sigma$-field. Let $X$ be a random variable and let $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function such that $\mathbb{E}(|X|), \mathbb{E}(|\varphi(X)|) < \infty$. Then

$$\varphi(\mathbb{E}(X|\,\mathcal{G})) \leq \mathbb{E}(\varphi(X)\,|\,\mathcal{G}).$$

**Proof.** Define

$$S = \left\{(a,b) \in \mathbb{R}^2 : \forall x \in \mathbb{R}\,[ax + b \leq \varphi(x)]\right\}.$$

Then

$$\varphi(x) = \sup_{(a,b)\in S} ax + b$$

Now, if $\varphi(x) \geq ax + b$ for all $x \in \mathbb{R}$, then

$$\varphi(X) \geq aX + b,$$

so that

$$\mathbb{E}(\varphi(X)\,|\,\mathcal{G}) \geq a\,\mathbb{E}(X|\,\mathcal{G}) + b.$$

Hence by taking supremum over $S$,

$$\mathbb{E}(\varphi(X)\,|\,\mathcal{G}) \geq \varphi(\mathbb{E}(X|\,\mathcal{G})).$$

**QED**

**Corollary 10.7.1.**

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-$\sigma$-field. Then

$$\|\mathbb{E}(X|\,\mathcal{G})\|_p \leq \|X\|_p$$

for $p \geq 1$.

**Proof.** It suffices to note that norms are convex, so that

$$\|\mathbb{E}(X|\,\mathcal{G})\|_p \leq \mathbb{E}\left(\|X\|_p\,|\,\mathcal{G}\right) = \|X\|_p.$$

**QED**

**Proposition 10.8.**
Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X$ be a $\mathrm{L}^2$ random variable on it. Then for any sub-$\sigma$-field $\mathcal{G} \subseteq \mathcal{F}$ and a random variable $Y$ on $(\Omega, \mathcal{G}, \mathbb{P}|_{\mathcal{G}})$,

$$\operatorname{cov}\left(X - \mathbb{E}\left(X \middle| \mathcal{G}\right), Y\right) = 0.$$

**Proof.** Observe that

$$\mathbb{E}\left(X - \mathbb{E}\left(X \middle| \mathcal{G}\right)\right) = \mathbb{E}(X) - \mathbb{E}(X) = 0,$$

so it remains to show that

$$\mathbb{E}\left(\left(X - \mathbb{E}\left(X \middle| \mathcal{G}\right)\right) Y\right) = 0.$$

Indeed,

$$\mathbb{E}\left(\left(X - \mathbb{E}\left(X \middle| \mathcal{G}\right)\right) Y\right) = \mathbb{E}\left(\mathbb{E}\left(\left(X - \mathbb{E}\left(X \middle| \mathcal{G}\right)\right) Y\right)\right) = \mathbb{E}\left(Y \mathbb{E}\left(X - \mathbb{E}\left(X \middle| \mathcal{G}\right) \middle| \mathcal{G}\right)\right) = \mathbb{E}\left(Y \cdot 0\right) = 0.$$

**QED**

**Corollary 10.8.1.**
Consider the setting of Proposition 10.8. $X, \mathbb{E}\left(X \middle| \mathcal{G}\right)$ are uncorrelated.

**Theorem 10.9.**
Let $X \in \mathrm{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ and let $Z \in \mathrm{L}^2(\Omega, \mathcal{G}, \mathbb{P}|_{\mathcal{G}})$ where $\mathcal{G} \subseteq \mathcal{F}$. Then

$$\mathbb{E}\left(\left(X - \mathbb{E}\left(X \middle| \mathcal{G}\right)\right)^2\right) \leq \mathbb{E}\left((X - Z)^2\right).$$

**Proof.** Observe that

$$\mathbb{E}\left((X - Z)^2\right) = \mathbb{E}\left(\left(\left(X - \mathbb{E}\left(X \middle| \mathcal{G}\right)\right) + \left(\mathbb{E}\left(X \middle| \mathcal{G}\right) - Z\right)\right)^2\right).$$

We have seen that $Z - \mathbb{E}\left(X \middle| \mathcal{G}\right), \mathbb{E}\left(X \middle| \mathcal{G}\right) - Z$ are uncorrelated. It follows that

$$\mathbb{E}\left((X - Z)^2\right) = \mathbb{E}\left(\left(X - \mathbb{E}\left(X \middle| \mathcal{G}\right)\right)^2\right) + \mathbb{E}\left(\left(\mathbb{E}\left(X \middle| \mathcal{G}\right) - Z\right)^2\right) \geq \mathbb{E}\left(\left(X - \mathbb{E}\left(X \middle| \mathcal{G}\right)\right)^2\right).$$

**QED**

**Corollary 10.9.1.** Wald's Identity
Let $X_1, X_2, \ldots$ be iid $\mathrm{L}^1$ random variables and let $N$ be a nonnegative integer-valued random variable with $\mathbb{E}(N) < \infty$ independent of $X_1, X_2, \ldots$. Then

$$\mathbb{E}\left(\sum_{n=1}^{N} X_n\right) = \mathbb{E}(X_1)\, \mathbb{E}(N).$$

**Proof.** Observe that

$$\mathbb{E}\left(\sum_{n=1}^{N} X_n\right) = \mathbb{E}\left(\mathbb{E}\left(\sum_{n=1}^{N} X_n \middle| N\right)\right) = \mathbb{E}\left(N \mathbb{E}(X_1)\right) = \mathbb{E}(N)\, \mathbb{E}(X_1),$$

since, when $N = k$,

$$\mathbb{E}\left(\sum_{n=1}^{N} X_n \middle| N = k\right) = \mathbb{E}\left(\sum_{n=1}^{k} X_n \middle| N = k\right) = \mathbb{E}\left(\sum_{n=1}^{k} X_n\right) = k\, \mathbb{E}(X_1) = N\, \mathbb{E}(X_1).$$

**QED**

**Corollary 10.9.2.** Eve's Law

Let $X, Y$ be random variables with $\mathbb{E}\left(X^2\right) < \infty$. Define ***conditional variance***

$$\operatorname{var}(X|Y) = \mathbb{E}\left((X - \mathbb{E}(X|Y))^2 | Y\right).$$

Then

$$\operatorname{var}(X) = \mathbb{E}(\operatorname{var}(X|Y)) + \operatorname{var}(\mathbb{E}(X|Y)).$$

**Proof.** Note that

$$\begin{aligned}\operatorname{var}(X) &= \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \mathbb{E}\left(((X - \mathbb{E}(X|Y)) + (\mathbb{E}(X|Y) - \mathbb{E}(X)))^2\right)\\ &= \mathbb{E}\left((X - \mathbb{E}(X|Y))^2\right) + \mathbb{E}\left((\mathbb{E}(X|Y) - \mathbb{E}(X))^2\right) = \mathbb{E}\left(\mathbb{E}\left((X - \mathbb{E}(X|Y))^2 | Y\right)\right) + \mathbb{E}\left((\mathbb{E}(X|Y) - \mathbb{E}(\mathbb{E}(X|Y)))^2\right)\\ &= \mathbb{E}(\operatorname{var}(X|Y)) + \operatorname{var}(\mathbb{E}(X|Y)).\end{aligned}$$

**QED**

Let $X_1, X_2, \ldots$ be iid random variables andl let $N$ be a nonnegative integer-valued random variable independent of $X_1, X_2, \ldots$. We would like to know about the distribution of $Y = \sum_{n=1}^{N} SX_n$.

Def'n 10.2. **Generating Function** of a Nonnegative Random Variable

The ***generating function*** of a nonnegative integer-valued random variable $Z$ is defined as

$$g_Z(t) = \mathbb{E}\left(t^Z\right) = \sum_{n=0}^{\infty} \mathbb{P}(Z = n)\, t^n.$$

Consider the characteristic function of $Y$:

$$\varphi_Y(t) = \mathbb{E}\left(e^{itY}\right) = \mathbb{E}\left(e^{it\sum_{n=1}^{N} X_n}\right) = \mathbb{E}\left(\mathbb{E}\left(e^{it\sum_{n=1}^{N} X_n} | N\right)\right), \qquad \forall t \in \mathbb{R}.$$

When $N = k$,

$$\mathbb{E}\left(e^{it\sum_{n=1}^{N} X_n} | N = k\right) = \mathbb{E}\left(e^{it\sum_{n=1}^{k} X_n} | N = k\right) = \mathbb{E}\left(e^{it\sum_{n=1}^{k} X_n}\right) = \mathbb{E}\left(e^{itX_1}\right)^k = \varphi_{X_1}(t)^k, \qquad \forall t \in \mathbb{R}.$$

It follows that

$$\varphi_Y(t) = \mathbb{E}\left(\varphi_{X_1}(t)^N\right) = \varphi_N\left(\varphi_{X_1}(t)\right), \qquad \forall t \in \mathbb{R}.$$

**Example 10.1.**

Suppose $X_1, X_2, \ldots \overset{\text{iid}}{\sim} \operatorname{EXP}(\lambda)$ and let $N \sim \operatorname{GEO}(p)$, where $N$ is independent of $X_1, X_2, \ldots$. Consider

$$Y = \sum_{n=1}^{N} X_n.$$

Observe that

$$\varphi_{X_1}(t) = \frac{\lambda}{\lambda - it}, g_N(t) = \frac{pt}{1 - (1-p)t}, \qquad \forall t \in \mathbb{R}.$$

This means

$$\varphi_Y(t) = \varphi_N\left(\varphi_{X_1}(t)\right) = \frac{p\frac{\lambda}{\lambda - it}}{1 - (1-p)\frac{\lambda}{\lambda - it}} = \frac{p\lambda}{p\lambda - it}, \qquad \forall t \in \mathbb{R},$$

which is the characteristic function of $\operatorname{EXP}(p\lambda)$. It follows that $Y \sim \operatorname{EXP}(p\lambda)$.