Department of Economics, University of Warwick
Mathematics for Economics
Lecture 9: Probability

Andrés Carvajal, with revisions by Peter Hammond

September, 2011; revised 2014 and 2016
Outline

Measures and Integrals

Measurable Spaces

Measure Spaces

Lebesgue Integration

Kolmogorov’s Definition of Probability

Cumulative Distribution and Density Functions

Expected Values

Limit Theorems

Convergence Results
Power Sets

Fix an abstract set $S \neq \emptyset$.

In case it is finite, its **cardinality**, denoted by $\#S$, is the number of distinct elements of $S$.

The **power set** of $S$ is the family $\mathcal{P}(S) := \{ T \mid T \subseteq S \}$ of all subsets of $S$.

**Exercise**

*Show that the mapping*

$$\mathcal{P}(S) \ni T \mapsto f(T) \in \{0, 1\}^S := \{ (x_s)_{s \in S} \mid \forall s \in S : x_s \in \{0, 1\} \}$$

*defined by*

$$f(T)_s = 1_T(s) = \begin{cases} 
1 & \text{if } s \in T \\
0 & \text{if } s \notin T 
\end{cases}$$

*is a bijection.*
Constructing a Bijection

Proof.
Evidently the mapping $T \mapsto f(T) = \langle 1_T(s) \rangle_{s \in S} \in \{0, 1\}^S$ is defined for every $T \subseteq S$ — i.e., for every $T \in \mathcal{P}(S)$.

Conversely, for each $\langle x_s \rangle_{s \in S} \in \{0, 1\}^S$, there is a unique set

$$T(\langle x_s \rangle_{s \in S}) = \{ s \in S \mid x_s = 1 \}$$

Now, given this $\langle x_s \rangle_{s \in S}$, for each $r \in S$, one has

$$1_{\{s \in S \mid x_s = 1\}}(r) = \begin{cases} 1 & \text{if } x_r = 1 \\ 0 & \text{if } x_r = 0 \end{cases}$$

So $f(T(\langle x_s \rangle_{s \in S})) = \langle 1_{\{s \in S \mid x_s = 1\}}(r) \rangle_{r \in S} = \langle x_r \rangle_{r \in S}$. This implies that $T(\langle x_s \rangle_{s \in S}) = f^{-1}(\langle x_s \rangle_{s \in S})$. $\square$
Counting a Finite Power Set

The existence of the bijection $\mathcal{P}(S) \ni T \mapsto f(T) \in \{0, 1\}^S$ proves that when $S$ is finite, then the cardinality of the power set is $\#\mathcal{P}(S) = 2^\#S$.

This helps explain why the power set $\mathcal{P}(S)$ is often denoted by $2^S$. 
Boolean Algebras, Sigma-Algebras, and Measurable Spaces

The family $\mathcal{A} \subseteq \mathcal{P}(S)$ is a **Boolean algebra** on $S$ just in case

1. $\emptyset \in \mathcal{A}$;
2. $A \in \mathcal{A}$ implies $S \setminus A \in \mathcal{A}$;
3. if $A, B$ lie in $\mathcal{A}$, then $A \cup B \in \mathcal{A}$.

The family $\Sigma \subseteq \mathcal{P}(S)$ is a **$\sigma$-algebra** just in case it is a Boolean algebra with the stronger property:

Whenever $\{A_n\}_{n=1}^{\infty}$ is a countably infinite family of sets in $\Sigma$, then $\bigcup_{n=1}^{\infty} A_n \in \Sigma$.

The pair $(S, \Sigma)$ is a **measurable space** just in case $\Sigma$ is a $\sigma$-algebra.
Exercise on Boolean Algebras and Sigma-Algebras

Exercise

1. Let $\mathcal{A}$ be a Boolean algebra on $S$.
   Prove that if $A, B \in \mathcal{A}$, then $A \cap B \in \mathcal{A}$.

2. Let $\Sigma$ be a $\sigma$-algebra on $S$.
   Prove that if $\{A_n\}_{n=1}^{\infty}$ is a countably infinite family of sets in $\Sigma$, then $\bigcap_{n=1}^{\infty} A_n \in \Sigma$.

Hint

1. For part 1, use de Morgan's laws

   \[
   S \setminus (A \cap B) = (S \setminus A) \cup (S \setminus B) \\
   S \setminus (A \cup B) = (S \setminus A) \cap (S \setminus B)
   \]

2. For part 2, use the infinite extension of de Morgan's laws:

   \[
   S \setminus (\bigcap_{n=1}^{\infty} A_n) = \bigcup_{n=1}^{\infty} (S \setminus A_n); \\
   S \setminus (\bigcup_{n=1}^{\infty} A_n) = \bigcap_{n=1}^{\infty} (S \setminus A_n)
   \]
Generating a Sigma-Algebra

Theorem

Let \( \{ \Sigma_i \mid i \in I \} \) be any indexed family of \( \sigma \)-algebras.

Then the intersection \( \bigcap \Sigma_i := \bigcap_{i \in I} \Sigma_i \) is also a \( \sigma \)-algebra.

Proof left as an exercise.

Let \( X \) be a space, and \( \mathcal{F} \subset 2^X \) any family of subsets.

Since \( 2^X \) is obviously a \( \sigma \)-algebra, there exists a non-empty set \( S(\mathcal{F}) \) of \( \sigma \)-algebras that include \( \mathcal{F} \).

Let \( \sigma(\mathcal{F}) \) denote the intersection \( \bigcap \{ \Sigma \mid \Sigma \in S(\mathcal{F}) \} \); it is the smallest \( \sigma \)-algebra that includes \( \mathcal{F} \).

Exercise

Let \( X \) be any uncountably infinite set, and let \( \mathcal{F} := \{ \{ x \} \mid x \in X \} \) denote the family of all singleton subsets of \( X \).

Show that \( \sigma(\mathcal{F}) \) consists of all subsets of \( X \) that are either countable, or co-countable (i.e., have a countable complement).
Topological Spaces

Given a set $X$, a topology $\mathcal{T}$ on $X$ is a family of open subsets $U \subseteq X$ satisfying:

1. $\emptyset \in \mathcal{T}$ and $X \in \mathcal{T}$;
2. if $U, V \in \mathcal{T}$, then $U \cap V \in \mathcal{T}$;
3. if $\{U_\alpha \mid \alpha \in A\}$ is any family of open sets in $\mathcal{T}$, then the union $\bigcup_{\alpha \in A} U_\alpha \in \mathcal{T}$.

Thus, finite intersections and arbitrary unions of open sets are open.

A topological space $(X, \mathcal{T})$ is any set $X$ together with a topology $\mathcal{T}$ that consists of all the open subsets of $X$. 
The Metric Topology

Let \((X, d)\) be any metric space.

The open ball of radius \(r\) centred at \(x\) is the set

\[B_r(x) := \{y \in X \mid d(x, y) < r\}\]

The metric topology \(\mathcal{T}_d\) of \((X, d)\) is the smallest topology that includes the entire family \(\{B_r(x) \mid x \in X \& r > 0\}\) of all open balls in \(X\).
Borel Sigma-Algebra

Let \((X, \mathcal{T})\) be any topological space.
Its Borel \(\sigma\)-algebra is defined as \(\sigma(\mathcal{T})\) — i.e., the smallest \(\sigma\)-algebra containing every open set of \(X\).

Suppose the topological space is a metric space \((X, d)\) with its metric topology \(\mathcal{T}_d\).
Then the Borel \(\sigma\)-algebra is generated by all the open balls \(B_r(x) := \{x' \in X \mid d(x, x') < r\}\) in \(X\).
For the case of the real line when \(X = \mathbb{R}\), its Borel \(\sigma\)-algebra is generated by all the open intervals of \(\mathbb{R}\).
Outline

Measures and Integrals

Measurable Spaces

Measure Spaces

Lebesgue Integration

Kolmogorov’s Definition of Probability

Cumulative Distribution and Density Functions

Expected Values

Limit Theorems

Convergence Results
Finitely Additive Set Functions

Let \( \bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\} = [-\infty, +\infty] \) denote the extended real line which, at each end, has an endpoint added at infinity.

Let \( \bar{\mathbb{R}}^+ := \mathbb{R}^+ \cup \{+\infty\} = [0, +\infty] \) be the non-negative part of \( \bar{\mathbb{R}} \).

Any family \( \mathcal{F} \) of subsets \( A \subseteq X \) is said to be pairwise disjoint just in case \( A \cap B = \emptyset \) whenever \( A, B \in \mathcal{F} \) with \( A \neq B \).

Definition

Let \((X, \Sigma)\) be a measurable space.

A mapping \( \mu : \Sigma \to \bar{\mathbb{R}}^+ \) is said to be a set function.

The set function \( \mu : \Sigma \to \bar{\mathbb{R}}^+ \) is said to be additive (or finitely additive) just in case, for any pair \( \{A, B\} \) of disjoint sets in \( \Sigma \), one has \( \mu(A \cup B) = \mu(A) + \mu(B) \).
Implications of Finite Additivity

Lemma
If the set function $\mu : \Sigma \to \bar{\mathbb{R}}_+$ is finitely additive, then $\mu(\emptyset) = 0$.

Proof.
For any non-empty $A \in \Sigma$, the sets $A$ and $\emptyset$ are disjoint.
Additivity implies that $\mu(A) = \mu(A \cup \emptyset) = \mu(A) + \mu(\emptyset)$, so $\mu(\emptyset) = 0$.

Exercise
For any finite collection $\{A_n\}_{n=1}^k$ of pairwise disjoint sets in $\Sigma$, prove by induction on $k$ that finite additivity implies

$$\mu \left( \bigcup_{n=1}^k A_n \right) = \sum_{n=1}^k \mu(A_n)$$
Definition
The set function $\mu : \Sigma \rightarrow \bar{\mathbb{R}}_+$ on a measurable space $(X, \Sigma)$ is said to be $\sigma$-additive or countably additive just in case, for any countable collection $\{A_n\}_{n=1}^\infty$ of pairwise disjoint sets in $\Sigma$, one has

$$\mu \left( \bigcup_{n=1}^\infty A_n \right) = \sum_{n=1}^\infty \mu(A_n)$$

A measure on a measurable space $(X, \Sigma)$ is a countably additive set function $\mu : \Sigma \rightarrow \bar{\mathbb{R}}_+$. 
Measure Space

A **measure space** is a triple \((X, \Sigma, \mu)\) where

1. \(\Sigma\) is a \(\sigma\)-algebra on \(X\);
2. \(\mu\) is a measure on the measurable space \((X, \Sigma)\).

**Example**

A prominent example of a measure space is \((\mathbb{R}, \mathcal{B}, \ell)\) where:

1. \(\mathcal{B}\) is the Borel \(\sigma\)-algebra induced by the open sets of the real line \(\mathbb{R}\);
2. the measure \(\ell(J)\) of any interval \(J \subset \mathbb{R}\) is its length, defined by \(\ell([a, b]) = \ell((a, b)) = \ell((a, b]) = \ell((a, b)) = b - a\);
3. \(\ell\) is extended to all of \(\mathcal{B}\) to satisfy countable additivity (it can be shown that this extension is unique).

**Exercise**

*Prove that if \(S \subset \mathbb{R}\) is countable, then \(S \in \mathcal{B}\) and \(\ell(S) = 0\).*
Lebesgue Measurable Subsets of the Real Line

A set $N \subset \mathbb{R}$ is null just in case there exists a Borel subset $B \in \mathcal{B}$ with $\ell(B) = 0$ such that $N \subset B$.

This is possible for some non-Borel subsets of $\mathbb{R}$. Let $\mathcal{N}$ denote the family of null subsets of $\mathbb{R}$. These null sets can be used to generate the Lebesgue $\sigma$-algebra of Lebesgue measurable sets, which is $\sigma(\mathcal{B} \cup \mathcal{N})$.

The symmetric difference of any two sets $S$ and $B$ is defined as the set

$$S \triangle B := (S \setminus B) \cup (B \setminus S) = (S \cup B) \setminus (S \cap B)$$

of elements $s$ that belong to one of the two sets, but not to both. One can show that $S \in \sigma(\mathcal{B} \cup \mathcal{N})$ if and only if there exists a Borel set $B \in \mathcal{B}$ such that $S \triangle B \in \mathcal{N}$ — i.e., $S$ differs from a Borel set only by a null set.
There is a well-defined function \( \lambda : \sigma(B \cup \mathcal{N}) \rightarrow \bar{R}_+ \) that satisfies \( \lambda(S) := \ell(B) \) whenever \( S \triangle B \in \mathcal{N} \).

Moreover, one can prove that the function \( S \mapsto \lambda(S) \) is countably additive.

This makes \( \lambda \) a measure, called the Lebesgue measure.

The associated measure space \( (\mathbb{R}, \sigma(B \cup \mathcal{N}), \lambda) \) is called the Lebesgue real line.
Outline

Measures and Integrals
  Measurable Spaces
  Measure Spaces

Lebesgue Integration

Kolmogorov’s Definition of Probability

Cumulative Distribution and Density Functions

Expected Values

Limit Theorems

Convergence Results
Measurable Partitions

Let \((X, \Sigma, \mu)\) be a measure space.

Given any set \(E \in \Sigma\), the indicator function of \(E\) is defined by

\[
X \ni x \mapsto 1_E(x) := \begin{cases} 
1 & \text{if } x \in E \\
0 & \text{if } x \notin E
\end{cases}
\]

The finite or countably infinite collection \(\{E_k| k \in K\}\) of pairwise disjoint measurable sets \(E_k \in \Sigma\)
is a measurable partition of \(X\) just in case \(\bigcup_{k \in K} E_k = X\).
Simple Functions

The function $f : X \to \mathbb{R}$ is simple just in case there exist a measurable partition $\{E_k|k \in K\}$ of $X$ together with a corresponding collection $(a_k)_{k \in K}$ of real constants such that $f(x) = \sum_{k \in K} a_k 1_{E_k}(x)$.

Note that the range $f(X) := \{y \in \mathbb{R} | \exists x \in X : y = f(x)\}$ of this step function is the precisely the set $\{a_k|k \in K\}$ of real constants.

Let $\mathcal{F}(X, \Sigma)$ denote the set of all simple functions on the measurable space $(X, \Sigma)$; in fact it is a real vector space.
Integrating Simple Functions

Given a function $f : X \mapsto \mathbb{R}$, whenever possible we want to define the integral $\int_X f(x) \, d\mu = \int_X f(x) \, \mu(dx)$.

The simple function $f(x) = \sum_{k \in K} a_k 1_{E_k}(x)$ is $\mu$-integrable just in case one has $\sum_{k \in K} |a_k| \mu(E_k) < +\infty$.

In particular, when $K$ is infinite, this requires the infinite series $\sum_{k \in K} a_k \mu(E_k)$ to be absolutely convergent.

Then we can define the integral $\int_X f(x) \, \mu(dx)$ of the $\mu$-integrable simple function $f(x) = \sum_{k \in K} a_k 1_{E_k}(x)$ as the real number $\sum_{k \in K} a_k \mu(E_k)$. 
Given the measure space \((X, \Sigma, \mu)\), the function \(f : X \rightarrow \mathbb{R}\) is measurable just in case, for every Borel set \(B \subset \mathbb{R}\), its inverse image satisfies

\[
f^{-1}(B) := \{ x \in X \mid f(x) \in B \} \in \Sigma
\]

Note that we have defined a simple function to be measurable.

Given any function \(f : X \rightarrow \mathbb{R}\), define the two sets

\[
\mathcal{F}^*(f; X, \Sigma) := \{ f^* \in \mathcal{F}(f; X, \Sigma) \mid \forall x \in X : f^*(x) \geq f(x) \}
\]

\[
\mathcal{F}_*(f; X, \Sigma) := \{ f_* \in \mathcal{F}(f; X, \Sigma) \mid \forall x \in X : f_*(x) \leq f(x) \}
\]

of simple functions that are respectively upper or lower bounds for the function \(f\).
Upper and Lower Integrals

The integral \( \int_X f^*(x) \mu(dx) \)
of each simple function \( f^* \in \mathcal{F}^*(f; X, \Sigma) \),
is an over-estimate of the true integral of \( f \).

But the integral \( \int_X f_*(x) \mu(dx) \)
of each simple function \( f_* \in \mathcal{F}_*(f; X, \Sigma) \),
is an under-estimate of the true integral of \( f \).

Define the \textbf{upper integral} and \textbf{lower integral} of \( f \) as, respectively

\[
I^*(f) := \inf_{f^* \in \mathcal{F}^*(f; X, \Sigma)} \int_X f^*(x) \mu(dx)
\]
and \( I_*(f) := \sup_{f_* \in \mathcal{F}_*(f; X, \Sigma)} \int_X f_*(x) \mu(dx) \)

Of course, in case \( f \) is itself a simple function,
one has \( I^*(f) = I_*(f) = \int_X f(x) \mu(dx) \).
Integration

Definition
The function $f : X \rightarrow \mathbb{R}$ is integrable just in case:

1. $f$ is measurable;
2. the upper integral $I^*(|f|)$ of the function $x \mapsto |f(x)|$ is defined (because $|f|$ is bounded above by an integrable simple function).

Theorem
The function $f : X \rightarrow \mathbb{R}$ is integrable if and only if:

1. $x \mapsto |f(x)|$ is bounded above by an integrable simple function;
2. and the upper and lower integrals $I^*(f)$ and $I_*(f)$ are equal.

So if $f : X \rightarrow \mathbb{R}$ is integrable, then we can define its integral $\int_X f(x) \, \mu(dx)$ as the common value of $I^*(f)$ and $I_*(f)$. 
Outline

Measures and Integrals
  Measurable Spaces
  Measure Spaces

Lebesgue Integration

Kolmogorov's Definition of Probability

Cumulative Distribution and Density Functions

Expected Values

Limit Theorems

Convergence Results
Probability Measure and Probability Space

Fix a measurable space \((S, \Sigma)\),
where \(S\) is a set of unknown states of the world.

Then \(\Sigma\) is a \(\sigma\)-algebra of unknown events.

A probability measure on \((S, \Sigma)\) is a measure \(\mathbb{P} : \Sigma \to \mathbb{R}_+\)
satisfying the requirement that \(\mathbb{P}(S) = 1\).

Countable additivity implies that \(\mathbb{P}(E) + \mathbb{P}(E^c) = 1\)
for every event \(E \in \Sigma\), where \(E^c := S \setminus E\).

It follows that \(\mathbb{P}(E) \in [0, 1]\) for every \(E \in \Sigma\).
Properties of Probability

Theorem

Let \((S, \Sigma, \mathbb{P})\) be a probability space.

Then the following hold for all \(\Sigma\)-measurable sets \(E, E'\) etc.

1. \(\mathbb{P}(E) \leq 1\) and \(\mathbb{P}(S \setminus E) = 1 - \mathbb{P}(E)\);

2. \(\mathbb{P}(E \setminus E') = \mathbb{P}(E) - \mathbb{P}(E \cap E')\) and
   \(\mathbb{P}(E \cup E') = \mathbb{P}(E) + \mathbb{P}(E') - \mathbb{P}(E \cap E')\);

3. for every partition \(\{E_n\}_{n=1}^m\) of \(S\) into \(m\) pairwise disjoint \(\Sigma\)-measurable sets, one has \(\mathbb{P}(E) = \sum_{n=1}^m \mathbb{P}(E \cap E_n)\);

4. \(\mathbb{P}(E \cap E') \geq \mathbb{P}(E) + \mathbb{P}(E') - 1\).

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

Proof.

We leave the routine proof as an exercise.
Two Limiting Properties

Theorem

Let \((S, \Sigma, \mathbb{P})\) be a probability space, and \((E_n)_{n=1}^{N}\) an infinite sequence of \(\Sigma\)-measurable sets.

1. If \(E_n \subseteq E_{n+1}\) for all \(n \in \mathbb{N}\),
   then \(\mathbb{P}(\bigcup_{n=1}^{\infty} E_n) = \lim_{n \to \infty} \mathbb{P}(E_n)\).

2. If \(E_n \supseteq E_{n+1}\) for all \(n \in \mathbb{N}\),
   then \(\mathbb{P}(\bigcap_{n=1}^{\infty} E_n) = \lim_{n \to \infty} \mathbb{P}(E_n)\).
Proving the Two Limiting Properties

Proof.

1. Because $E_n \subseteq E_{n+1}$ for all $n \in \mathbb{N}$, one has

$$E_n = E_1 \cup \left[ \bigcup_{k=2}^{n} (E_k \setminus E_{k-1}) \right]$$
and

$$\bigcup_{n=1}^{\infty} E_n = E_1 \cup \left[ \bigcup_{k=2}^{\infty} (E_k \setminus E_{k-1}) \right]$$

where the sets $E_1$ and \{\(E_k \setminus E_{k-1}\) | \(k = 2, 3, \ldots\)\} are all pairwise disjoint. Hence

$$\mathbb{P}(E_n) = \mathbb{P}(E_1) + \sum_{k=2}^{n} \mathbb{P}(E_k \setminus E_{k-1})$$
$$\mathbb{P}(\bigcup_{n=1}^{\infty} E_n) = \mathbb{P}(E_1) + \sum_{k=2}^{\infty} \mathbb{P}(E_k \setminus E_{k-1})$$

$$= \lim_{n \to \infty} \left[ \mathbb{P}(E_1) + \sum_{k=2}^{n} \mathbb{P}(E_k \setminus E_{k-1}) \right]$$
$$= \lim_{n \to \infty} \mathbb{P}(E_n)$$

2. Apply part 1 to the complements of the sets $E_n$. 

□
Conditional Probability: First Definition

Let \( E^* \in \Sigma \) be such that \( \mathbb{P}(E^*) > 0 \).

The conditional probability measure on \( E^* \) is the mapping

\[
\Sigma \ni E \mapsto \mathbb{P}(E|E^*) := \frac{\mathbb{P}(E \cap E^*)}{\mathbb{P}(E^*)} \in [0, 1]
\]

The triple \((E^*, \Sigma(E^*), \mathbb{P}(\cdot|E^*))\) with

\[
\Sigma(E^*) := \{ E \cap E^* \mid E \in \Sigma \} = \{ E \in \Sigma \mid E \subseteq E^* \}
\]

is then a conditional probability space given the event \( E^* \).
Conditional Probability: Two Key Properties

Theorem
Provided that $\mathbb{P}(E) \in (0,1)$, one has

$$\mathbb{P}(E') = \mathbb{P}(E)\mathbb{P}(E'|E) + (1 - \mathbb{P}(E))\mathbb{P}(E'|E^c)$$

Theorem
Let $(E_k)_{k=1}^n$ be any finite list of sets in $\Sigma$. Provided that $\mathbb{P}(\cap_{k=1}^{n-1} E_k) > 0$, one has

$$\mathbb{P}(\cap_{k=1}^n E_k) = \mathbb{P}(E_1)\mathbb{P}(E_2|E_1)\mathbb{P}(E_3|E_1 \cap E_2) \ldots \mathbb{P}(E_n|\cap_{k=1}^{n-1} E_k)$$
Independence

The finite or countably infinite family \( \{E_k\}_{k \in K} \) of events in \( \Sigma \) is:

- **pairwise independent** if \( \mathbb{P}(E \cap E') = \mathbb{P}(E) \mathbb{P}(E') \) whenever \( E \neq E' \);
- **independent** if for any finite subfamily \( \{E_k\}_{k=1}^n \), one has \( \mathbb{P}(\cap_{k=1}^n E_k) = \prod_{k=1}^n \mathbb{P}(E_k) \).

Exercise

Let \( S \) be the set \( \{1, 2, 3, 4, 5, 6, 7, 8, 9\} \), and \( \mathbb{P} \) the probability measure on \( 2^S \) satisfying \( \mathbb{P}(\{s\}) = 1/9 \) for all \( s \in S \).

Consider the three events

\[
E_1 = \{1, 2, 7\}, \quad E_2 = \{3, 4, 7\} \quad \text{and} \quad E_3 = \{5, 6, 7\}
\]

Are these event pairwise independent? Are they independent?

Exercise

Prove that if \( \{E, E'\} \) is independent, then so is \( \{E^c, E'\} \).
Random Variable

Definition

- The function $X : S \rightarrow \mathbb{R}$ is $\Sigma$-measurable just in case for every $x \in \mathbb{R}$ one has

$$X^{-1}(-\infty, x) := \{s \in S \mid X(s) \leq x\} \in \Sigma$$

- A random variable (with values in $\mathbb{R}$) is a $\Sigma$-measurable function $X : S \rightarrow \mathbb{R}$.

- The distribution function or cumulative distribution function (cdf) of $X$ is the mapping $F_X : \mathbb{R} \rightarrow [0, 1]$ defined by

$$x \mapsto F_X(x) = \mathbb{P}(\{s \in S \mid X(s) \leq x\})$$
Outline

Measures and Integrals
  Measurable Spaces
  Measure Spaces

Lebesgue Integration

Kolmogorov’s Definition of Probability

Cumulative Distribution and Density Functions

Expected Values

Limit Theorems

Convergence Results
Properties of Distribution Functions, I

**Theorem**

The CDF of any random variable \( s \mapsto X(s) \) satisfies:

1. \( \lim_{x \to -\infty} F_X(x) = 0 \) and \( \lim_{x \to +\infty} F_X(x) = 1 \).
2. \( x \geq x' \) implies \( F_X(x) \geq F_X(x') \).
3. \( \lim_{h \downarrow 0} F_X(x + h) = F_X(x) \).
4. \( \mathbb{P}(\{s \in S \mid X(s) > x\}) = 1 - F_X(x) \).
5. \( \mathbb{P}(\{s \in S : x < X(s) \leq x'\}) = F_X(x') - F_X(x) \) whenever \( x < x' \),
6. \( \mathbb{P}(\{s \in S : X(s) = x\}) = F_X(x) - \lim_{h \uparrow 0} F_X(x + h) \).

CDFs are sometimes said to be càdlàg, which is a French acronym for *continue à droite, limite à gauche* (continuous on the right, limit on the left).
Properties of Distribution Functions, II

Definition
A continuity point of the CDF $F_X : \mathbb{R} \to [0, 1]$ is an $\bar{x} \in \mathbb{R}$ at which the mapping $x \mapsto F_X(x)$ is continuous.

Is it always true that $\lim_{h \uparrow 0} F_X(x + h) = F_X(x)$?

Exercise
Let $F_X : \mathbb{R} \to [0, 1]$ be the CDF of any random variable $S \ni s \mapsto X(s) \to \mathbb{R}$, and $\bar{x} \in \mathbb{R}$ any point.

Prove that the following three conditions are equivalent:

1. $\bar{x}$ is a continuity point of $F_X$;
2. $\mathbb{P}(\{s \in S \mid X(s) = \bar{x}\}) = 0$;
3. $\lim_{h \uparrow 0} F_X(x + h) = F_X(x)$. 
Continuous Random Variable

Definition

- A random variable $S \ni s \mapsto X(s) \to \mathbb{R}$ is
  1. continuous if $x \mapsto F_X(x)$ is continuous;
  2. absolutely continuous if there exists a density function $\mathbb{R} \ni x \mapsto f_X(x) \to \mathbb{R}_+$ such that $F_X(x) = \int_{-\infty}^{x} f_X(u)du$ for all $x \in \mathbb{R}$.

- The support of the random variable $S \ni s \mapsto X(s) \to \mathbb{R}$ is the closure of the set on which $F_X$ is strictly increasing.

Example

The uniform distribution on a closed interval $[a, b]$ of $\mathbb{R}$ has density function $f$ and distribution function $F$ given by

$$f_X(x) := \frac{1}{b-a} \mathbf{1}_{[a, b]}(x)$$

and

$$F_X(x) := \begin{cases} 
0 & \text{if } x < a \\
\frac{x-a}{b-a} & \text{if } x \in [a, b] \\
1 & \text{if } x > b
\end{cases}$$
The Normal or Gaussian Distribution

Example

The standard normal distribution on $\mathbb{R}$ has density function $f$ given by

$$f_X(x) := ke^{-\frac{1}{2}x^2}$$

where $k := \frac{1}{\sqrt{2\pi}}$ is chosen so that $\int_{-\infty}^{+\infty} ke^{-\frac{1}{2}x^2} dx = 1$.

Its mean and variance are

$$\int_{-\infty}^{+\infty} kxe^{-\frac{1}{2}x^2} dx = \lim_{a \to \infty} \int_{-a}^{+a} kxe^{-\frac{1}{2}x^2} dx$$

$$= \lim_{a \to \infty} \left[ - \int_{0}^{a} kxe^{-\frac{1}{2}x^2} dx + \int_{0}^{a} kxe^{-\frac{1}{2}x^2} dx \right]$$

$$= 0$$

$$\int_{-\infty}^{+\infty} kx^2 e^{-\frac{1}{2}x^2} dx = 1$$
The Gaussian Integral, $I$

Define $I(a) := \int_{-a}^{+a} e^{-\frac{1}{2}x^2} \, dx$ for each $a \in \mathbb{R}$. Then

$$[I(a)]^2 = \left( \int_{-a}^{+a} e^{-\frac{1}{2}x^2} \, dx \right) \left( \int_{-a}^{+a} e^{-\frac{1}{2}y^2} \, dy \right)$$

$$= \int_{-a}^{+a} \left( \int_{-a}^{+a} e^{-\frac{1}{2}y^2} \, dy \right) e^{-\frac{1}{2}x^2} \, dx$$

$$= \int_{-a}^{+a} \int_{-a}^{+a} e^{-\frac{1}{2}x^2} e^{-\frac{1}{2}y^2} \, dxdy$$

$$= \int_{S(a)} e^{-\frac{1}{2}(x^2+y^2)} \, dxdy$$

where $S(a) := [-a, a]^2$ denotes the Cartesian product of the line interval $[-a, a]$ with itself.

Thus, $S(a)$ is the solid square subset of $\mathbb{R}^2$ that is centred at the origin and has sides of length $2a$. 
The Gaussian Integral, II

Let \( D(b) := \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq b^2\} \)

denote the disk of radius \( b \) centred at the origin.

Consider the transformation \((r, \theta) \mapsto (x, y) = (r \cos \theta, r \sin \theta)\)

from polar to Cartesian coordinates.

The Jacobian matrix of this transformation is

\[
\begin{vmatrix}
\frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\
\frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta}
\end{vmatrix}
= \begin{vmatrix}
\cos \theta & -r \sin \theta \\
\sin \theta & r \cos \theta
\end{vmatrix}
= r(\cos^2 \theta + \sin^2 \theta) = r
\]

It follows that changing to polar coordinates

in the double integral \( J(b) = \int_{D(b)} e^{-\frac{1}{2}(x^2+y^2)} \, dx \, dy \)

transforms it to

\[
J(b) = \int_0^b \int_0^{2\pi} re^{-\frac{1}{2}r^2} \, dr \, d\theta
= \left( \int_0^b re^{-\frac{1}{2}r^2} \, dr \right) \left( \int_0^{2\pi} 1 \, d\theta \right)
= \left( \left| \int_0^b (-e^{-\frac{1}{2}r^2}) \right| \right) 2\pi
= 2\pi \left( 1 - e^{-\frac{1}{2}b^2} \right)
\]
The Gaussian Integral, III

Note that \( D(a) \subset S(a) \subset D(a\sqrt{2}) \) and so

\[
\int_{D(a)} e^{-\frac{1}{2}(x^2+y^2)} \, dx \, dy \leq \int_{S(a)} e^{-\frac{1}{2}(x^2+y^2)} \, dx \, dy \leq \int_{D(a\sqrt{2})} e^{-\frac{1}{2}(x^2+y^2)} \, dx \, dy
\]

It follows that

\[
J(a) \leq [I(a)]^2 = \int_{S(a)} e^{-\frac{1}{2}(x^2+y^2)} \, dx \, dy \leq J(a\sqrt{2})
\]

and so \(2\pi(1 - e^{-\frac{1}{2}a^2}) \leq [I(a)]^2 \leq 2\pi(1 - e^{-a^2}).\)

Taking limits as \(a \to \infty\) one has \(2\pi(1 - e^{-\frac{1}{2}a^2}) \to 2\pi\)
and also \(2\pi(1 - e^{-a^2}) \to 2\pi\), implying that \([I(a)]^2 \to 2\pi\).

**Theorem**

The Gaussian integral \(\int_{-\infty}^{+\infty} e^{-\frac{1}{2}x^2} \, dx\) equals \(\sqrt{2\pi}\).
Outline

Measures and Integrals
  Measurable Spaces
  Measure Spaces

Lebesgue Integration

Kolmogorov’s Definition of Probability

Cumulative Distribution and Density Functions

Expected Values

Limit Theorems

Convergence Results
Expectation

Let $g : \mathbb{R} \to \mathbb{R}$ be any Borel function, and $x \mapsto f_X(x)$ the density function of the random variable $X$. Whenever the integral $\int_{-\infty}^{\infty} |g(x)| f_X(x) \, dx$ exists, the expectation of $g \circ X$ is defined as

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) \, dx$$

Theorem

Let $g_1, g_2 : \mathbb{R} \to \mathbb{R}$ and $a, b, c \in \mathbb{R}$. Then:

1. $\mathbb{E}(ag_1(X) + bg_2(X) + c) = a\mathbb{E}(g_1(X)) + b\mathbb{E}(g_2(X)) + c$.
2. If $g_1 \geq 0$, then $\mathbb{E}(g_1(X)) \geq 0$.
3. If $g_1 \geq g_2$, then $\mathbb{E}(g_1(X)) \geq \mathbb{E}(g_2(X))$. 
Chebychev’s Inequality

Theorem
For any random variable $S \ni s \mapsto X(s) \in Z$, fix any measurable function $g : Z \to \mathbb{R}_+$ with $\mathbb{E}[g(X(s))] < +\infty$. Then for all $r > 0$ one has $\mathbb{P}(g(X) \geq r) \leq \frac{1}{r} \mathbb{E}[g(X)]$.

Proof.
The two indicator functions $s \mapsto 1_{g(X) \geq r}(s)$ and $s \mapsto 1_{g(X) < r}(s)$ satisfy $1_{g(X) \geq r}(s) + 1_{g(X) < r}(s) = 1$ for all $s \in S$.

Because $g(X(s)) \geq 0$ for all $s \in S$, one has

$$
\mathbb{E}[g(X)] = \mathbb{E}[\{1_{g(X) \geq r}(s) + 1_{g(X) < r}(s)\} g(X(s))] = \mathbb{E}[1_{g(X) \geq r}(s) g(X(s))] + \mathbb{E}[1_{g(X) < r}(s) g(X(s))] \geq r \mathbb{E}[1_{g(X) \geq r}(s)] = r \mathbb{P}(g(X) \geq r)
$$

Dividing by $r$ implies that $\frac{1}{r} \mathbb{E}[g(X)] \geq \mathbb{P}(g(X) \geq r)$.
Moments and Central Moments

For a random variable $X$ and any $k \in \mathbb{N}$:

- its $k^{th}$ (noncentral) moment is $\mathbb{E}[X^k]$;
- its $k^{th}$ central moment is $\mathbb{E}[(X - \mathbb{E}[X])^k]$, assuming that $\mathbb{E}[X]$ exists in $\mathbb{R}$;
- its variance, $\text{Var} X$, is its second central moment.
Odd Central Moments of the Gaussian Distribution

Let \( m_n := \int_{-\infty}^{+\infty} kx^n e^{-\frac{1}{2}x^2} \, dx \) denote the \( n \)th central moment of the standard normal distribution.

When \( n \) is odd, one has \((-x)^n = -x^n\), so

\[
m_n = \int_{-\infty}^{+\infty} kx^n e^{-\frac{1}{2}x^2} \, dx
\]

\[
= \lim_{a \to \infty} \int_{-a}^{0} kx^n e^{-\frac{1}{2}x^2} \, dx + \lim_{a \to \infty} \int_{0}^{+a} kx^n e^{-\frac{1}{2}x^2} \, dx
\]

\[
= -\lim_{a \to \infty} \int_{0}^{a} kx^n e^{-\frac{1}{2}x^2} \, dx + \lim_{a \to \infty} \int_{0}^{a} kx^n e^{-\frac{1}{2}x^2} \, dx
\]

\[
= 0
\]
Even Central Moments of the Gaussian Distribution

Now suppose \( n = 2r \), where \( r \in \mathbb{N} \).

Because \( \frac{d}{dx} e^{-\frac{1}{2}x^2} = -x e^{-\frac{1}{2}x^2} \), integrating by parts gives

\[
\int_{-a}^{+a} kx^n e^{-\frac{1}{2}x^2} \, dx = -\int_{-a}^{+a} kx^{n-1} \left( \frac{d}{dx} e^{-\frac{1}{2}x^2} \right) \, dx
\]

\[
= - \bigg|_{-a}^{+a} kx^{n-1} e^{-\frac{1}{2}x^2} + \int_{-a}^{+a} k(n-1)x^{n-2}e^{-\frac{1}{2}x^2} \, dx
\]

\[
= -k[a^{n-1} - (-a)^{n-1}]e^{-\frac{1}{2}a^2} + \int_{-a}^{+a} k(n-1)x^{n-2}e^{-\frac{1}{2}x^2} \, dx
\]

Taking the limit as \( a \to \infty \), one obtains \( m_n = (n-1)m_{n-2} \).

Note that \( m_0 = 1 \), so when \( n \) is an even integer \( 2r \), one has

\[
m_{2r} = (2r - 1)(2r - 3) \cdots 5 \cdot 3 \cdot 1
\]

\[
= \frac{2r(2r - 1)(2r - 2)(2r - 3) \cdots 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{2r(2r - 2)(2r - 4) \cdots 6 \cdot 4 \cdot 2} = \frac{(2r)!}{2^r r!}
\]
Multiple Random Variables

Let \( S \ni s \mapsto \mathbf{X}(s) = (X_n(s))_{n=1}^N \)
be an \( N \)-dimensional vector of random variables
defined on the probability space \( (S, \Sigma, \mathbb{P}) \).

- Its joint distribution function is the mapping defined by

\[
\mathbb{R}^N \ni \mathbf{x} \mapsto F_{\mathbf{X}}(\mathbf{x}) := \mathbb{P}(\{s \in S \mid \mathbf{X}(s) \leq \mathbf{x}\})
\]

- The random vector \( \mathbf{X} \) is absolutely continuous
just in case there exists a density function \( f_{\mathbf{X}} : \mathbb{R}^N \to \mathbb{R}_+ \)
such that

\[
F_{\mathbf{X}}(\mathbf{x}) = \int_{\mathbf{u} \leq \mathbf{x}} f_{\mathbf{X}}(\mathbf{u}) \, d\mathbf{u} \quad \text{for all } \mathbf{x} \in \mathbb{R}^N
\]
Independent Random Variables

Let \( \mathbf{X} \) be an \( N \)-dimensional vector valued random variable.

- If \( \mathbf{X} \) is absolutely continuous,
  the marginal density \( \mathbb{R} \ni x \mapsto f_{X_n}(x) \) of its \( n \)th component \( X_n \) is defined as the \( N-1 \)-dimensional iterated integral

  \[
  f_{X_n}(x) = \int \cdots \int f_{\mathbf{X}}(x_1, \ldots, x_{n-1}, x, x_{n+1}, \ldots, x_N) \, dx_1 \ldots dx_N
  \]

  in which every variable except \( X_n \) gets “integrated out”.

- The \( N \) components of \( \mathbf{X} \) are independent just in case \( f_{\mathbf{X}} = \prod_{n=1}^{N} f_{X_n} \).

- The infinite sequence \( (X_n)_{n=1}^{\infty} \) of random variables is independent just in case every finite subsequence \( (X_n)_{n \in K} (K \text{ finite}) \) is independent.
Expectations

Let $\mathbf{X}$ be an $N$-dimensional vector valued random variable, and $g : \mathbb{R}^N \to \mathbb{R}$ a measurable function.

The expectation of $g(\mathbf{X})$ is defined as the $N$-dimensional integral

$$\mathbb{E}[g(\mathbf{X})] := \int_{\mathbb{R}^N} g(\mathbf{u}) f_{\mathbf{X}}(\mathbf{u}) \, d\mathbf{u}$$

Theorem

If the collection $(X_n)_{n=1}^N$ of random variables is independent, then

$$\mathbb{E} \left[ \prod_{n=1}^N X_n \right] = \prod_{n=1}^N \mathbb{E}(X_n)$$

Exercise

Prove that if the pair $(X_1, X_2)$ of r.v.s is independent, then its covariance satisfies

$$\text{Cov}(X_1, X_2) := \mathbb{E}[(X_1 - \mathbb{E}[X_1])(X_2 - \mathbb{E}[X_2])] = 0$$
Marginal and Conditional Density

Fix the pair \((X_1, X_2)\) of random variables.

- The **marginal density** of \(X_1\) is
  
  \[
  f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{(X_1,X_2)}(x_1, x_2) \, dx_2.
  \]

- At points \(x_1\) where \(f_{X_1}(x_1) > 0\),
  the **conditional density** of \(X_2\) given that \(X_1 = x_1\) is
  
  \[
  f_{X_2 | X_1}(x_2 | x_1) = \frac{f_{(X_1,X_2)}(x_1, x_2)}{f_{X_1}(x_1)}
  \]

**Theorem**

*If the pair \((X_1, X_2)\) is independent and \(f_{X_1}(x_1) > 0\), then*

\[
 f_{X_2 | X_1}(x_2 | x_1) = f_{X_2}(x_2)
\]
Conditional Expectations

Fix the pair \((X_1, X_2)\) of random variables.

- The conditional expectation of \(g(X_2)\) given that \(X_1 = x_1\) is

\[
\mathbb{E}[g(X_2)|X_1 = x_1] = \int_{-\infty}^{\infty} g(x_2) f_{X_2|X_1}(x_2|x_1) dx_2.
\]

- Given any measurable function \((x_1, x_2) \mapsto g(x_1, x_2)\), the law of iterated expectations states that

\[
\mathbb{E}_{f(x_1, x_2)}[g((X_1, X_2)(s))] = \mathbb{E}_{f_X_1}[\mathbb{E}_{f_{X_2|X_1}}[g((X_1, X_2)(s))]]
\]

Proof.

\[
\mathbb{E}_{f(x_1, x_2)}[g] = \int_{\mathbb{R}^2} g(x_1, x_2) f_{(X_1, X_2)}(x_1, x_2) dx_1 dx_2 \\
= \int_{\mathbb{R}} \left[ \int_{\mathbb{R}} g(x_1, x_2) f_{X_2|X_1}(x_2|x_1) dx_2 \right] f_{X_1}(x_1) dx_1 \\
= \mathbb{E}_{f_{X_1}}[\mathbb{E}_{f_{X_2|X_1}} g(x_1, x_2)]
\]
Outline

Measures and Integrals
  Measurable Spaces
  Measure Spaces
Lebesgue Integration
Kolmogorov’s Definition of Probability
Cumulative Distribution and Density Functions
Expected Values

Limit Theorems
Convergence Results
Convergence of Random Variables

The sequence \((X_n)_{n=1}^{\infty}\) of random variables:

- **converges in probability to** \(X\) (written as \(X_n \overset{p}{\to} X\))
  just in case, for all \(\epsilon > 0\), one has
  \[
  \lim_{n \to \infty} \mathbb{P}(|X_n - X| < \epsilon) = 1.
  \]

- **converges in distribution to** \(X\) (written as \(X_n \overset{d}{\to} X\))
  just in case, for all \(x\) at which \(F_X\) is continuous,
  \[
  \lim_{n \to \infty} F_{X_n}(x) = F_X(x)
  \]
Definition of Weak Convergence

**Definition**
Let \((X, \Sigma, P)\) be any probability space.

Then a **continuity set** of \((X, \Sigma, P)\) is any set \(B \in \Sigma\) whose boundary \(\partial B\) satisfies \(P(\partial B) = 0\).

**Definition**
Let \((X, d)\) be a metric space with its Borel -algebra \(\Sigma\).

A sequence \((P_n)_{n \in \mathbb{N}}\) of probability measures on the measurable space \((X, \Sigma)\) **converges weakly** to the probability measure \(P\), written \(P_n \Rightarrow P\), just in case \(P_n(B) \to P(B)\) as \(n \to \infty\) for any continuity set of \((X, \Sigma, P)\).
Portmanteau Theorem

Theorem

Let $P$ and $(P_n)_{n \in \mathbb{N}}$ be probability measures on the measurable space $(X, \Sigma)$.

Then $P_n \Rightarrow P$ if and only if:

1. for all bounded continuous functions $f : X \to \mathbb{R}$, one has:
   \[
   \int_X f(x) P_n(dx) \to \int_X f(x) P(dx)
   \]

2. \( \limsup_{n \to \infty} P_n(C) \leq P(C) \) for every closed subset \( C \subseteq X \);

3. \( \liminf_{n \to \infty} P_n(U) \geq P(U) \) for every open set \( U \subseteq X \).
Convergence of Distribution Functions

**Theorem**

Let $F$ and $(F_n)_{n \in \mathbb{N}}$ be cumulative distribution functions on $\mathbb{R}$ with associated probability measures $P$ and $(P_n)_{n \in \mathbb{N}}$ on the Lebesgue real line that satisfy

$$F(x) = P((\neg \infty, x]) \quad \text{and} \quad F_n(x) = P_n((\neg \infty, x]) \quad (n \in \mathbb{N})$$

on the measurable space $(X, \Sigma)$.

Then $P_n \Rightarrow P$ if and only if $F_n(x) \to F(x)$ for all $x$ at which $F$ is continuous.
The Law of Large Numbers

- The sequence \((X_n)_{n=1}^{\infty}\) of random variables is i.i.d.
  — i.e., independently and identically distributed
  — just in case
    1. it is independent, and
    2. for every Borel set \(D \subseteq \mathbb{R}\), one has \(P(X_n \in D) = P(X'_n \in D)\).

- The weak law of large numbers:
  Let \((X_n)_{n=1}^{\infty}\) be i.i.d. with \(E(X_n) = \mu\).
  Define the sequence
  \[
  (\bar{X}_n)_{n=1}^{\infty} := \left(\frac{1}{n} \sum_{k=1}^{n} X_k\right)_{n=1}^{\infty}
  \]
  of sample means. Then, \(\bar{X}_n \xrightarrow{p} \mu\).
The Meaning of Probability

Prove the following:

Let $\gamma = p(X \in \Omega) \in (0, 1)$.

Consider the following experiment:
“$n$ realizations of $X$ are taken independently.”

Let $G_n$ be the relative frequency with which a realization in $\Omega$ is obtained in the experiment.

Then, $G_n \xrightarrow{p} \gamma$. 
The Central Limit Theorem

- The central limit theorem:
  Let \((X_n)_{n=1}^\infty\) be i.i.d. random variables with \(\mathbb{E}(X_n) = \mu\) and \(V(X_n) = \sigma^2\). Then,

\[
\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{d} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}u^2} \, du.
\]
The Fundamental Theorems

Let \((X_n)_{n=1}^{\infty}\) be i.i.d., with \(E[X_n] = \mu\) and \(V(X_n) = \sigma^2\). Then:

- by the law of large numbers,
  \[
  \bar{X}_n \xrightarrow{p} \mu;
  \]

  so
  \[
  \bar{X}_n \xrightarrow{d} \mu;
  \]

- but by the central limit theorem,
  \[
  \frac{\bar{X}_n - \mu}{(\sigma/\sqrt{n})} \xrightarrow{d} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}u^2} \, du.
  \]
Concepts of Convergence, 1

Definition
Say that the sequence $X_n$ of random variables converges almost surely or with probability 1 or strongly towards $X$ just in case

$$\liminf_{n \to \infty} \mathbb{P}(\{\omega \in \Omega \mid |X_n(\omega) - X(\omega)| < \epsilon\}) = 1$$

Hence, the values of $X_n$ approach those of $X$, in the sense that the event that $X_n(\omega)$ does not converge to $X(\omega)$ has probability 0.

Almost sure convergence is often denoted by $X_n \xrightarrow{P-a.s.} X$, with the letters a.s. over the arrow that indicates convergence.
For generic random elements $X_n$ on a metric space $(S, d)$, almost sure convergence is defined similarly, replacing the absolute value $|X_n(\omega) - X(\omega)|$ by the distance $d(X_n(\omega), X(\omega))$.

Almost sure convergence implies convergence in probability, and *a fortiori* convergence in distribution.

It is the notion of convergence used in the strong law of large numbers.
The Strong law

Definition
The strong law of large numbers (or SLLN) states that the sample average $\bar{X}_n$ converges almost surely to the expected value $\mu = \mathbb{E}X$.

It is this law (rather than the weak LLN) that justifies the intuitive interpretation of the expected value of a random variable as its “long-term average when sampling repeatedly.”
Differences Between the Weak and Strong Laws

The weak law states that for a specified large $n$, the average $\bar{X}_n$ is likely to be near $\mu$.

This leaves open the possibility that $|\bar{X}_n - \mu| \geq \epsilon$ happens an infinite number of times, although at infrequent intervals.

The strong law shows that this almost surely will not occur.

In particular, it implies that with probability 1, for any $\epsilon > 0$ there exists $n_\epsilon$ such that $|\bar{X}_n - \mu| < \epsilon$ holds for all $n > n_\epsilon$. 
Moment-Generating Functions

Definition
The $n$th moment about the origin is defined as $m_n := \mathbb{E}[X^n]$. This may not exist for large $n$ unless the random variable $X$ is essentially bounded, meaning that there exists an upper bound $\bar{x}$ such that $\mathbb{P}(\{ \omega \in \Omega \mid |X(\omega) \leq \bar{x}\}) = 1$.

Definition
The moment-generating function of a random variable $X$ is
$$\mathbb{R} \ni t \mapsto M_X(t) := \mathbb{E}[e^{tX}]$$
wherever this expectation exists.
At $t = 0$, of course, $M_X(0) = 1$.
For $t \neq 0$, however, unless $X$ is essentially bounded above, the expectation typically may not exist because $e^{tX}$ can be unbounded.
The Gaussian Case

For a normal or Gaussian distribution \( N(\mu, \sigma^2) \), even though the random variable is unbounded, the tails of the distribution vanish quickly enough so that the moment generating function exists and is given by

\[
M(t; \mu, \sigma^2) = \mathbb{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx
\]

Now make the substitution \( y = (x - \mu - \sigma^2 t)/\sigma \), implying that \( dx = \sigma dy \) and that

\[
tx - \frac{(x - \mu)^2}{2\sigma^2} = -\frac{1}{2}y^2 + \mu t + \frac{1}{2}\sigma^2 t^2
\]

This transforms the integral to

\[
M(t; \mu, \sigma^2) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} e^{\mu t + \frac{1}{2}\sigma^2 t^2} \, dy = e^{\mu t + \frac{1}{2}\sigma^2 t^2}
\]
Note that
\[
e^{tX} = 1 + tX + \frac{t^2X^2}{2!} + \frac{t^3X^3}{3!} + \cdots + \frac{t^nX^n}{n!} + \cdots = \sum_{n=0}^{\infty} \frac{t^nX^n}{n!}
\]

Taking the expectation term by term and then using the definition of the moments of the distribution, one obtains

\[
M_X(t) = \mathbb{E}[e^{tX}]
\]
\[
= 1 + t\mathbb{E}[X] + \frac{t^2}{2!}\mathbb{E}[X^2] + \frac{t^3}{3!}\mathbb{E}[X^3] + \cdots + \frac{t^n}{n!}\mathbb{E}[X^n] + \cdots
\]
\[
= 1 + tm_1 + \frac{t^2}{2!}m_2 + \frac{t^3}{3!}m_3 + \cdots + \frac{t^k}{k!}m_k + \cdots
\]
\[
= \sum_{k=0}^{\infty} \frac{t^k}{k!}m_k
\]
Derivatives of the Moment-Generating Function

Suppose we find the \( n \)th derivative with respect to \( t \) of \( M_X(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} m_k \).

Note that \( \frac{d^n}{dt^n} t^k = k(k-1)(k-2)\ldots(k-n+1) = \frac{k!}{(k-n)!} t^{k-n} \) as is easily proved by induction on \( n \).

So differentiating term by term \( n \) times, one obtains

\[
M_X^{(n)}(t) = \mathbb{E} \left[ \frac{d^n}{dt^n} e^{tX} \right] = \sum_{k=n}^{\infty} \frac{k!}{(k-n)!} \frac{t^{k-n}}{k!} m_k
\]

Putting \( t = 0 \) yields the equality \( M_X^{(n)}(0) = \frac{t^0}{0!} m_n = m_n \).

In this sense, the moment-generating function does “exponentially generate” the moments of the probability distribution.
Definition of Characteristic Functions

The moment-generating function may not exist because the expectation need not converge absolutely.

By contrast, the expectation of the bounded function $e^{itX}$ always lies in the unit circle of the complex plane $\mathbb{C}$.

So the characteristic function that we are about to introduce always exists, which makes it more useful in many contexts.

Definition

For a scalar random variable $X$ with CDF $x \mapsto F_X(x)$, the characteristic function is defined as the (complex) expected value of $e^{itX} = \cos tX + i \sin tX$, where $i = \sqrt{-1}$ is the imaginary unit, and $t \in \mathbb{R}$ is the argument of the characteristic function:

$$\mathbb{R} \ni t \mapsto \phi_X(t) = \mathbb{E}e^{itX} = \int_{-\infty}^{+\infty} e^{itx} dF_X(x) \in \mathbb{C}$$
Gaussian Case

Consider a normally distributed random variable $X$ with mean $\mu$ and variance $\sigma^2$.

Its characteristic function can be found by replacing $t$ by $it$ in the expression for the moment

$$M(t; \mu, \sigma^2) = \mathbb{E}[e^{tX}] = e^{\mu t + \frac{1}{2} \sigma^2 t^2}$$

Recalling that $(it)^2 = -t^2$, the result is

$$\phi(t; \mu, \sigma^2) = \int_{-\infty}^{\infty} e^{itx} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} (x-\mu)^2 / \sigma^2} dx = e^{i\mu t - \frac{1}{2} \sigma^2 t^2}$$

In the standard normal or $N(0, 1)$ case, when $\mu = 0$ and $\sigma^2 = 1$, one has $\phi(t; 0, 1) = e^{-\frac{1}{2} t^2}$. 
Use of Characteristic Functions

Characteristic functions can be used
to give superficially simple proofs
of both the LLN and the classical central limit theorems.

The following merely sketches the argument.

For much more careful detail,
see Richard M. Dudley’s major text, *Real Analysis and Probability*.

A key tool is Lévy’s continuity theorem.

For a sequence of random variables,
this connects convergence in distribution
to pointwise convergence of their characteristic functions.
Statement of Lévy’s Continuity Theorem

Theorem
Suppose \((X_n)_{n=1}^{\infty}\) is a sequence of random variables, not necessarily sharing a common probability space, with the corresponding sequence

\[ \mathbb{R} \ni t \mapsto \varphi_n(t) = \mathbb{E} e^{itX_n} \in \mathbb{C} \quad (n \in \mathbb{N}) \]

of complex-valued characteristic functions.

If \(X_n\) converges in distribution to the random variable \(X\), then \(t \mapsto \varphi_n(t)\) converges pointwise to \(t \mapsto \varphi(t) = \mathbb{E} e^{itX}\), the characteristic function of \(X\).

Conversely, if \(t \mapsto \varphi_n(t)\) converges pointwise to a function \(t \mapsto \varphi(t)\) which is continuous at \(t = 0\), then \(t \mapsto \varphi(t)\) is the characteristic function \(\mathbb{E} e^{itX}\) of a random variable \(X\), and \(X_n\) converges in distribution to \(X\).
Linear Approximation to the Characteristic Function

Suppose that the random variable $X$ has a mean $\mu_X := \mathbb{E}X = \int_{-\infty}^{\infty} x dF(x)$.

One can then differentiate within the expectation to obtain

$$\frac{d}{dt} \mathbb{E}e^{itX} = \mathbb{E}\left[\frac{d}{dt} e^{itX}\right] = \mathbb{E}[iXe^{itX}]$$

Consider the linear approximation

$$\mathbb{E}e^{ihX} = 1 + i[\mu + \xi(h)]h \quad \text{where} \quad \xi(h) := (\mathbb{E}e^{ihX} - 1 - ih\mu)/h$$

By l’Hôpital’s rule, one has

$$\lim_{h \to 0} \xi(h) = "0/0" = \lim_{h \to 0} (\mathbb{E}[iXe^{ihX}] - i\mu)/1 = \mathbb{E}[iX] - i\mu = 0$$
Quadratic Approximation to the Characteristic Function

Next, suppose that the random variable $X$ has not only a mean $\mu_X := \int_{-\infty}^{\infty} x dF(x)$, but also a variance $\sigma^2_X := \int_{-\infty}^{\infty} (x - \mu)^2 dF(x)$.

One can then differentiate twice within the expectation to obtain

$$\frac{d^2}{dt^2} \mathbb{E} e^{itX} = \mathbb{E} \left[ \frac{d^2}{dt^2} e^{itX} \right] = \mathbb{E}[(iX)^2 e^{itX}] = -\mathbb{E}[X^2 e^{itX}]$$

Consider the quadratic approximation

$$\mathbb{E} e^{ihX} = 1 + i\mu h - \frac{1}{2} \left[ \sigma^2 + \mu^2 + \eta(h) \right] h^2$$

where $\eta(h) := (1/h^2) \left[ \mathbb{E} e^{ihX} - 1 - ih\mu \right] + \frac{1}{2} (\sigma^2 + \mu^2)$.

Applying l'Hôpital's rule twice, one has

$$\lim_{h \to 0} \frac{1}{h^2} \left[ \mathbb{E} e^{ihX} - 1 - ih\mu \right] = "0/0" = \lim_{h \to 0} \frac{1}{2h} \left( \mathbb{E}[iX e^{ihX}] - i\mu \right)$$

$$= "0/0" = \lim_{h \to 0} \frac{1}{2} \mathbb{E}[(iX)^2 e^{ihX}] = -\frac{1}{2} \mathbb{E} X^2 = -\frac{1}{2} (\sigma^2 + \mu^2)$$

implying that $\eta(h) \to 0$ as $h \to 0$. 
A Useful Lemma

Lemma

Suppose that $\mathbb{R} \ni h \mapsto \zeta(h) \in \mathbb{C}$ satisfies $\zeta(h) \to 0$ as $h \to 0$.

Then for all $z \in \mathbb{C}$, one has $\{1 + \frac{1}{n}[z + \zeta(1/n)]\}^n \to e^z$ as $n \to \infty$.

For a sketch proof, first one can show that

$$\lim_{n \to \infty} \left\{1 + \frac{1}{n}[z + \zeta(1/n)]\right\}^n = \lim_{n \to \infty} \left(1 + \frac{1}{n}z\right)^n$$

Second, in case $z \in \mathbb{R}$, putting $h = 1/n$ and taking logs gives

$$\ln \left[\lim_{n \to \infty} \left(1 + \frac{1}{n}z\right)^n\right] = \ln \left[\lim_{h \to 0} (1 + hz)^{1/h}\right]$$

$$= \lim_{h \to 0} \frac{1}{h} \left[\ln(1 + hz) - \ln 1\right] = \frac{d}{dh} \ln(1 + hz)\bigg|_{h=0} = z$$

implying that $(1 + \frac{1}{n}z)^n \to e^z$ as $n \to \infty$.

Dealing with the case when $z$ is complex is more tricky.
Consider now any infinite sequence $X_1, X_2, \ldots$ of observations of IID random variables drawn from a common CDF $F(x)$ on $\mathbb{R}$, with common characteristic function $t \mapsto \varphi_X(t) = \mathbb{E}[e^{itX}]$.

For each $n \in \mathbb{N}$, let $\bar{X}_n := \frac{1}{n} \sum_{j=1}^{n} X_j$ denote the random variable whose value is the sample mean of the first $n$ observations.

This sample mean has its own characteristic function

$$\varphi_{\bar{X}_n}(t) := \mathbb{E}[e^{it\bar{X}_n}] = \mathbb{E} \left[ \prod_{j=1}^{n} e^{itX_j/n} \right]$$

Then

$$\varphi_{\bar{X}_n}(t) = \prod_{j=1}^{n} \mathbb{E}[e^{itX_j/n}] = \left( \mathbb{E}[e^{itX/n}] \right)^n$$

because the random variables $X_j$ are respectively independently and identically distributed.
Suppose we take the linear approximation

$$\mathbb{E}e^{ihX} = 1 + i[\mu + \xi(h)]h \quad \text{where} \quad \xi(h) \to 0 \quad \text{as} \quad h \to 0$$

and replace $h$ by $t/n$ to obtain

$$\mathbb{E}[e^{it\bar{X}_n}] = \{1 + (it/n)[\mu + \xi(t/n)]\}^n$$

Because $\xi(t/n) \to 0$ as $n \to \infty$ and so $h = t/n \to 0$, one has

$$\lim_{n \to \infty} \{1 + \frac{1}{n}it[\mu + \xi(t/n)]\}^n = \lim_{n \to \infty} (1 + \frac{1}{n}it\mu)^n = e^{it\mu} = \mathbb{E}[e^{itY}]$$

where $\mathbb{E}[e^{itY}]$ is the characteristic function of a degenerate random variable $Y$ which is equal to $\mu$ with probability 1.

Using the Lévy theorem, it follows that the distribution of $\bar{X}_n$ converges to this degenerate distribution, implying that $\bar{X}_n$ converges to $\mu$ in probability.
Sketch Proof of the Classical CLT, 1

For each \( j \in \mathbb{N} \), let \( Z_j \) denote the standardized value \((X_j - \mu)/\sigma\) of \( X_j \), defined to have the property that \( \mathbb{E}Z_j = 0 \) and \( \mathbb{E}Z_j^2 = 1 \).

Now define \( \bar{Z}_n := \sum_{j=1}^{n} \frac{Z_j}{\sqrt{n}} \), which is called the standardized mean because:

1. linearity implies that \( \mathbb{E}\bar{Z}_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \mathbb{E}Z_j = 0 \);
2. independence implies that \( \mathbb{E}\bar{Z}_n^2 = \frac{1}{n} \sum_{j=1}^{n} \mathbb{E}Z_j^2 = 1 \).

Putting \( \mu = 0 \) and \( \sigma^2 = 0 \) in the quadratic approximation

\[
\mathbb{E}e^{ihX} = 1 + i\mu h - \frac{1}{2} [\sigma^2 + \mu^2 + \eta(h)] h^2
\]

implies \( \mathbb{E}e^{ihZ} = 1 - \frac{1}{2} [1 + \eta(h)] h^2 \) where \( \eta(h) \to 0 \) as \( h \to 0 \).

Replacing \( hX \) by \( tZ_j/\sqrt{n} \) in this quadratic approximation yields

\[
\mathbb{E}[e^{itZ_j/\sqrt{n}}] = 1 - \frac{1}{2} \frac{t^2}{n} [1 + \eta(t/n)]
\]
Sketch Proof of the Classical CLT, II

Now independence implies that

$$\mathbb{E}[e^{it\bar{Z}_n}] = \mathbb{E}\left[\exp\left(it\frac{1}{\sqrt{n}} \sum_{j=1}^{n} Z_j\right)\right] = \prod_{j=1}^{n} \mathbb{E}[e^{itZ_j / \sqrt{n}}]$$

Hence, another careful limiting argument shows that

$$\mathbb{E}[e^{it\bar{Z}_n}] = \left\{1 - \frac{1}{2} \frac{t^2}{n} [1 + \eta(t/n)]\right\}^n \to e^{-\frac{1}{2}t^2} \text{ as } n \to \infty$$

But we showed that this limit $e^{-\frac{1}{2}t^2}$ is precisely the characteristic function of a standard normal distribution $N(0, 1)$.

So the central limit theorem follows from the Lévy continuity theorem, which confirms that the convergence of characteristic functions implies convergence in distribution.