Random Process
Lecture 1. Fundamentals of Probability

Husheng Li

Min Kao Department of Electrical Engineering and Computer Science
University of Tennessee, Knoxville

Spring, 2016
Outline

1. Syllabus

2. Measure Theory

3. Probability Spaces

4. Convergence

5. Conditioning
Syllabus

- Homework: 4 problems each week; hand in your homework after one week. 40%
- Midterm and Final Exames: 60%
Goal of This Course

- Understand the underlying math for probability and random processes — measure theory
- Understand the basic concepts of probabilities such as distribution, expectation and convergence.
- Understand various types of random processes
- Know the applications in engineering
Outline

- Lecture 1. Measure Theory and Probability
- Lecture 2. Martingales
- Lecture 3. Markov Processes
- Lecture 4. Poisson Processes
- Lecture 5. Brownian Motion
- Lecture 6. Levy Processes
- Lecture 7. Stationary Process and Power Spectrum
Outline

1. Syllabus
2. Measure Theory
3. Probability Spaces
4. Convergence
5. Conditioning
A nonempty collection $\mathcal{E}$ of subsets of set $E$ is called an algebra provided that it is closed under finite unions and complements.

It is called a $\sigma$-algebra if it is closed under complements and countable unions:

$$A \in \mathcal{E} \rightarrow E/A \in \mathcal{E}.$$ 

and

$$A_1, A_2, ... \in \mathcal{E} \rightarrow \bigcup_n A_n \in \mathcal{E}.$$ 

Why $\sigma$-algebra?
Measurable Spaces and Functions

- A measurable space is a pair \((E, \mathcal{E})\) where \(E\) is a set and \(\mathcal{E}\) is a \(\sigma\)-algebra on \(E\).

- Product of measurable spaces \((E, \mathcal{E})\) and \((F, \mathcal{F})\): \((E \times F, \mathcal{E} \otimes \mathcal{F})\).

- A mapping \(f : E \rightarrow F\) is said to be measurable relative to \(\mathcal{E}\) and \(\mathcal{F}\) if \(f^{-1}(B) \in \mathcal{E}\) for any \(B \in \mathcal{F}\).
Indicators and Simple Functions

- Let $A \subset E$, its indicator $1_A$ is defined as
  \[ 1_A(x) = \begin{cases} 
    1, & \text{if } x \in A, \\
    0, & \text{otherwise}. 
  \end{cases} \]

- A function $f$ is called simple if
  \[ f = \sum_{i=1}^{n} a_i 1_{A_i} \]
Measures

- A measure $\mu$ on $(E, \mathcal{E})$ is a mapping from $\mathcal{E}$ to nonnegative real numbers such that
  \[ \mu(\emptyset) = 0, \]
  and
  \[ \mu(\bigcup_n A_n) = \sum_n \mu(A_n), \]
  for every disjoint sequence $\{A_n\}$.

- A measure space is a triplet $(E, \mathcal{E}, \mu)$.

- Almost everywhere: If a proposition holds for all but a negligible set of $x$ in $E$, then we say that it holds for almost every $x$, or almost everywhere.
Example of Measures

- Dirac measures.
- Counting measures.
- Discrete measures.
- Lebesgue measures.
Integration

- When $f$ is simple and positive. Then we define its integration as

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

- For positive measurable function $f$, we can use a series of simple functions to approach it $f_n \to f$. Then we define

$$
\mu f = \lim \mu f_n.
$$

- For generic measurable function $f$, we define

$$
\mu f = \mu(f^+) - \mu(f^-).
$$

- If $\mu f$ exists and is a real number, we say that $f$ is measurable.
Examples of Integrations

- Discrete measures.
- Discrete spaces.
- Lebesgue integrals.
- Integral of Dirichlet function.
Intuition of Lebesgue Integrals: Counting Moneys

Lebesgue integral and Riemann integral imply different approaches of counting money.
Convergence Theorems

- Monotone Convergence Theorem: Let \( f_n \) be increasing in \( \mathcal{E}_+ \), then
  \[
  \mu(\lim f_n) = \lim \mu f_n.
  \]

- Dominated Convergence Theorem: Suppose \( \{f_n\} \) is dominated by some integrable function \( g \). If \( \lim f_n \) exists, then it is integrable and
  \[
  \mu(\lim f_n) = \lim \mu f_n.
  \]

- Fatou’s Lemma: if \( \{f_n\} \in \mathcal{E}_+ \), then
  \[
  \mu(\lim \inf f_n) \leq \lim \inf \mu f_n.
  \]
Differentiation

Let $\mu$ and $\nu$ be measures on a measurable space. Then, $\nu$ is said to be absolutely continuous with respect to $\mu$, if for every set $A \in \mathcal{E}$

$$\mu(A) = 0 \Rightarrow \nu(A) = 0.$$ 

Radon-Nikodym Theorem: Suppose $\mu$ is $\sigma$-finite, and $\nu$ is absolutely continuous w.r.t. $\mu$. Then there exists a positive measurable function $p$ such that

$$\int \nu(dx)f(x) = \int \mu(dx)p(x)f(x).$$

We can denote $p$ by $d\nu/d\mu$ and call it Radon-Nikodym derivative.
Kernel

- For measurable spaces \((E, \mathcal{E})\) and \((F, \mathcal{F})\), \(K\), a mapping from \(E \times \mathcal{F}\) to \(R^+\), is called a kernel if
  - the mapping \(x \rightarrow K(x, B)\) is \(\mathcal{E}\)-measurable for every set \(B\) in \(\mathcal{F}\).
  - the mapping \(B \rightarrow K(x, B)\) is a measure on \((F, \mathcal{F})\) for every \(x\) in \(E\).
Outline

1 Syllabus
2 Measure Theory
3 Probability Spaces
4 Convergence
5 Conditioning
Let $(\Omega, \mathcal{H}, P)$ be a probability. $\Omega$ is called the sample space, its elements are called the outcomes. $\mathcal{H}$ is called the event space.

An event is said to be almost sure if its probability is 1.
Let \((E, \mathcal{E})\) be a measurable space. A mapping \(X : \Omega \to E\) is called a random variable if \(X^{-1}A\) is measurable in \(\mathcal{H}\).

**Distribution:** \(\mu(A) = P(X^{-1}A), \text{ for } A \in \mathcal{E}\).

**Independence:**

\[
P(X \in A, Y \in B) = P(X \in A)P(Y \in B).
\]
Stochastic Processes

- Let \((E, \mathcal{E})\) be a measurable space. Let \(T\) be an arbitrary set, countable or uncountable. For any \(t\) in \(T\), let \(X_t\) be a random variable taking values in \((E, \mathcal{E})\). Then, the collection \(\{X_t, t \in T\}\) is called a stochastic process with state space \((E, \mathcal{E})\).

- One can consider the random process \(X_t\) as a random variable taking values in the product space \((E^T, \mathcal{E}^T)\).

- The distribution of random process is determined by probability of finite samples:

\[
P(X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n).
\]
The expectation of a random variable $X$ is given by

$$EX = \int_{\Omega} P(dw)X(w).$$

If $X$ takes values in a measurable space $(E, \mathcal{E})$ and $\mu$ is the distribution of $X$, then we have

$$E[f(X)] = \mu f.$$
**$L^p$-Space**

- Define the $p$-norm of random variable $X$:
  \[
  \|X\|_p = (E|X|^p)^{1/p}.
  \]

  Let $L^p$ denote the collection of all real-valued random variables $X$ with $\|X\|_p < \infty$.

- Holder’s inequality: if $1/p + 1/q = 1/r$, then
  \[
  \|XY\|_r \leq \|X\|_p \|Y\|_q.
  \]

- Mikowski’s Inequality: for $p \geq 1$
  \[
  \|X + Y\|_p \leq \|X\|_p + \|Y\|_p.
  \]
Uniform Integrability

- A collection $K$ of real valued random variables is said to be uniformly integrable if
  \[
  \lim_{b \to \infty} = \sup_{X \in K} E[|X|1(|X| > b)].
  \]

- Necessary and sufficient condition for uniform integrability: for any $\epsilon > 0$ we can find a $\delta > 0$ such that for every event $H$
  \[
  P(H) \leq \delta \Rightarrow \sup_{X \in K} E|X|1_H \leq \epsilon
  \]
Information and Determinability

- We call the set \( \{ X^{-1} A, A \in \mathcal{E} \} \) the \( \sigma \)-algebra generated by \( X \). For a collection of random variables \( \{ X_t | t \in T \} \), we define the \( \sigma \)-algebra generated by \( \{ X_t \} \) as

\[
\sigma \{ X_t : t \in T \} = \bigvee_{t \in T} \sigma X_t.
\]

- A mapping \( V \) belongs to \( \sigma X \) if and only if \( V = f(X) \) for some deterministic function \( f \) in \( \mathcal{E} \).

- We can equate the information available from the realization of \( \{ X_t \} \) to the \( \sigma \)-algebra generated by \( \{ X_t \} \).
Filtrations

Let $T$ be a subset of $R$. For each $t \in T$, let $\mathcal{F}_t$ be a sub-$\sigma$-algebra of $\mathcal{H}$. The family $\mathcal{F} = \{\mathcal{F}_t : t \in T\}$ is called a filtration given that $\mathcal{F}_s \subset \mathcal{F}_t$ for $s < t$.

A typical filtration:

$$\mathcal{F}_t = \sigma\{X_s : s \leq t, t \in T\}.$$
Independence

Let $\mathcal{F}_1, \ldots, \mathcal{F}_n$ be sub-$\sigma$-algebra of $\mathcal{H}$. Then, $\{\mathcal{F}_1, \ldots, \mathcal{F}_n\}$ is called an independency if

$$E[V_1 \ldots V_n] = E[V_1] \ldots E[V_n],$$

for all positive random variables $V_1, \ldots, V_n$ in $\mathcal{F}_1, \ldots, \mathcal{F}_n$ respectively.

$\mathcal{F}_1, \ldots, \mathcal{F}_n, \ldots$ of $\mathcal{H}$ are independent if and only if $\mathcal{F}_1, \ldots, \mathcal{F}_n$ and $\mathcal{F}_{n+1}$ are independent.

Independent random variables: $\{X_t, t \in T\}$ is independent if $\sigma\{X_t, t \in T\}$ is an independency.
0-1 Laws

- Let $G_n$ be a sequence of sub-$\sigma$-algebra of $\mathcal{H}$. Define $\mathcal{F}_n = \bigvee_{m > n} G_m$ and the tail-$\sigma$-algebra $\mathcal{F} = \bigcap_n \mathcal{F}_n$.

- Kolmogorov 0-1 Law: Let $G_1, \ldots, G_n, \ldots$ be independent. Then $P(H)$ is either 0 or 1 for every event $H$ in the tail $\mathcal{F}$.

- Consider a permutation $p$ and $X = (X_1, X_2, \ldots)$. Then, $X \circ p = (X_{p(1)}, \ldots, X_{p(2)}, \ldots)$. A random variable $V$ in $\mathcal{F}_\infty$ is said to be permutation invariant if $V \circ p = V$.

- Hewit 0-1 Law: Suppose that $X_1, \ldots, X_n, \ldots$ are i.i.d. Then every permutation invariant event has probability 0 or 1.
Outline

1 Syllabus
2 Measure Theory
3 Probability Spaces
4 Convergence
5 Conditioning
Homework Problems


- Problem 1: Let $\mathcal{F}$ be a $\sigma$-algebra of subsets of $\Omega$. Suppose $B \in \mathcal{F}$. Show that $\mathcal{G} = \{A \cap B : A \in \mathcal{F}\}$ is a $\sigma$-algebra of subsets of $B$.

- Problem 2: Consider a random variable mapping from $(\Omega, \mathcal{H}, P)$ to $(E, \mathcal{E})$. Prove that the set generated by $X$, $\{X^{-1}(A) | A \in \mathcal{E}\}$ is a $\sigma$-algebra.

- Problem 3: Use both the Riemann integral and Lebesgue integral to evaluate $\int_0^1 1 \, dx$.

- Problem 4: Show that Kolmogorov 0-1 Law implies the law of large numbers.
Almost Sure Convergence

- A sequence \( \{X_n\} \) is said to be almost surely convergent if the numerical sequence \( \{X_n(w)\} \) is convergent for almost every \( w \).

- Borel-Cantelli’s Lemma: Let \( H_n \) be a sequence of events. Then
  \[
  \sum P(H_n) < \infty \Rightarrow \sum 1_{H_n} < \infty \text{almost surely}.
  \]
  If \( \sum_n P(|X_n - X| > \epsilon) < \infty \), then \( X_n \to X \) almost surely.
Convergence in Probability

- The sequence of random variables $\{X_n\}$ converges to $X$ in probability if for every $\epsilon > 0$ we have
  \[ \lim_{n} P(|X_n - X| > \epsilon) = 0. \]

- If $X_n$ converges almost surely, then it converges in probability.
- If $X_n$ converges in probability, then it has a subsequence converging almost surely.
Convergence in $L^p$

The sequence of random variables $\{X_n\}$ converges to $X$ in $L^p$ if every $X_n$ and $X$ are in $L^p$ and

$$\lim_{n} E|X_n - X|^p = 0.$$ 

The following are equivalent:

- $\{X_n\}$ converges in $L^1$
- It converges in probability and is uniformly integrable.
- It is Cauchy for convergence in $L^1$. 
Weak Convergence

- Let \( \{X_n\} \) has distributions \( \{\mu_n\} \). \( \{\mu_n\} \) is said to converge weakly to \( \mu \) if
  \[
  \lim \mu_n f = \mu f,
  \]
  for any \( f \) in \( C_b \). The sequence \( \{X_n\} \) is said to converge to \( X \) if \( \mu_n \) converges to \( \mu \).

- A sequence \( \mu_n \) is said to be tight if for every \( \epsilon > 0 \) there is a compact set \( K \) such that \( \mu_n(K) > 1 - \epsilon \) for all \( n \).

- Theorem: If \( \{\mu_n\} \) is tight then every subsequence of it has a further subsequence that is weakly convergent.
Law of Large Numbers

- For \(\{X_i\}\), define \(S_n = X_1 + ... + X_n\) and \(\bar{X}_n = \frac{1}{n}S_n\).

- Theorem: If \(\{X_n\}\) is pairwise i.i.d, then \(\bar{X}_n\) converges in \(L^2\), in probability and almost surely.

- Theorem: If \(\{X_i\}\) is pairwise independent and have the same distribution as a generic variable \(X\). Then, if \(EX\) exists, \(\bar{X}\) converges to \(EX\) almost surely.
Central Limit Theorem

Let \( \{X_i\} \) be i.i.d. Let \( S_n = \sum_{i=1}^{n} X_i \) and \( Z_n = \frac{S_n - nE[X_i]}{\sqrt{nVar[X_i]}} \). Then, \( Z_n \) converge to a standard Gaussian random variable in distribution.

The law of large numbers and the central limit theorem are the most important theorems in probably theory. There are many variants of these two fundamental theorems (e.g., what if the samples are not i.i.d?).
Outline

1. Syllabus
2. Measure Theory
3. Probability Spaces
4. Convergence
5. Conditioning
Conditional Expectations

Let $\mathcal{F}$ be a sub-$\sigma$-algebra of $\mathcal{H}$. The conditional expectation of $X$ given $\mathcal{F}$ denoted by $E_\mathcal{F}X$ is defined in two steps:

- For $X$ in $\mathcal{H}_+$, the conditional expectation is defined as any random variable $X$ that satisfies (1) $X$ belongs to $\mathcal{F}_+$ and $EVX = EVX$ for every $V$ in $\mathcal{F}_+$.
- For arbitrary $X$ in $\mathcal{H}$, we define

$$E_\mathcal{F}(X) = E_\mathcal{F}(X^+) - E_\mathcal{F}(X^-).$$
Let $\mathcal{F}$ and $\mathcal{G}$ be sub-$\sigma$-algebra of $\mathcal{H}$. Let $W$ and $X$ be random variables (in $\mathcal{H}$) such that $EX$ and $EWX$ exist. Then we have

- **Conditional determinism:** $W \in \mathcal{G} \Rightarrow E_{\mathcal{F}}W = WE_{\mathcal{F}}X$.
- **Repeated conditioning:** $\mathcal{F} \subset \mathcal{G} \Rightarrow E_{\mathcal{F}}E_{\mathcal{G}}X = E_{G}E_{\mathcal{F}}X = E_{\mathcal{F}}X$
Conditional Expectations
Given Random Variable

Recall $\sigma Y$ is the $\sigma$-algebra generated by random variable $Y$. Then, we define

$$E[X|Y] = E_{\sigma Y} X.$$
Let $\mathcal{F}$ be a sub-$\sigma$-algebra of $\mathcal{H}$. For each $H$ in $\mathcal{H}$,

$$P_{\mathcal{F}}H = E_{\mathcal{F}}1_{\mathcal{H}}$$

is called the conditional probability of $H$ given $\mathcal{F}$.

Let $Q(H)$ be a version of $P_{\mathcal{F}}H$ for $H$ in $\mathcal{H}$. Then, $Q$ is said to be a regular version of the conditional probability if $Q$ is a transition probability kernel from $(\Omega, \mathcal{F})$ into $(\Omega, \mathcal{H})$. 
Conditional Independence

Let \( \{ \mathcal{F}_i \} \) be sub-\( \sigma \)-algebras of \( \mathcal{H} \). Then they are said to be conditional independent given \( \mathcal{F} \) if

\[
E_{\mathcal{F}} V_1 \ldots V_n = E_{\mathcal{F}} V_1 \ldots E_{\mathcal{F}} V_n.
\]