## Contents

1. Set 1: Fundamentals of Probability  
   1.1. Experiments 6  
   1.2. Outcomes and the Sample Space 6  
   1.3. Events 6  
   1.4. Random Variables 9  
2. Set 2: Fundamentals of Probability, Continued 11  
   2.1. Probability Measure 11  
   2.2. Expectation 13  
3. Lecture 1: 08/25/03 16  
   3.1. Definition of Sample Space, Events and Random Variables 16  
   3.2. Sub-algebras 17  
4. Lecture 2: 08/27/03 18  
   4.1. Set Theory Notation 18  
   4.2. Random Variables 19  
5. Lecture 3: 09/03/2003 21  
   5.1. Random Variables 21  
   5.2. Properties of Random Variables 21  
   5.3. Measure 22  
   5.4. Expectation 23  
   6.1. Expectation 24  
   6.2. Distributions 25  
   6.3. Distribution Function 25  
   6.4. Density Functions 26  
7. Lecture 5: 09/10/2003 27  
   7.1. Elementary Hilbert Space Theory 27  
   7.2. Schwarz Inequality 27  
   7.3. Triangle Inequality 27  
   7.4. Convex Sets 28  
   7.5. Orthogonality 28  
8. Lecture 6: 09/15/03 29  
   8.1. Triangle Inequality 29  
   8.2. Notion of a Closed Set 29  
   8.3. Application 30
19. Lecture 17: 10/29/03

19.1. Random Processes

19.2. Wide Sense Stationary (WSS)

19.3. Example of Random Processes

20. Lecture 18: 11/03/03

20.1. Review

20.2. Brownian Motion

20.3. White Noise

21. Lecture 19: 11/05/03

21.1. White Noise

21.2. Independent Increment Hypothesis

21.3. Application of White Noise

21.4. Norton Model for Thermal Noise

22. Lecture 20: 11/10/03

22.1. Practical Interpretation

22.2. Connection between $R_X(z)$ and $S_X(\omega)$

22.3. Response of LTI Systems to WSS Stochastic Process

23. Lecture 21: 11/12/03

23.1. Linear Systems Response

24. Lecture 22: 11/17/03

24.1. Fact about Gaussian r.v.'s

24.2. Conditional Expectations

24.3. Poisson Process

24.4. Hazard or Failure Rates

25. Lecture 23: 11/19/03

25.1. Linear Estimation: Discrete Case

25.2. Linear Estimation: Continuous Case

26. Lecture 24: 11/24/03

26.1. Linear Filtering

26.2. Causal Wiener Filtering

27. Lecture 25: 11/26/03

27.1. Spectral Factorization

27.2. Optimal Linear Prediction

27.3. Smoothing

27.4. Markov Chains
1. Set 1: Fundamentals of Probability

1.1. Experiments. The most fundamental component in probability theory is the existence of a physical or abstract “experiment,” whose “outcome” is revealed when the experiment is completed. Probability theory aims to provide the tools that will enable us to assess the likelihood of an outcome, or more generally, the likelihood of any outcome from a collection of outcomes. Let us consider the following example:

Example 1. Shooting a single dart: Consider shooting a single dart at a target (board) represented by the unit closed disc, $D$, which is centered at the point $(0, 0)$. We write $D = \{ (x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1 \}$. Here, $\mathbb{R}$ denotes the set of real numbers (same as $(-\infty, \infty)$), and $\mathbb{R}^2$ is the set of all points in the plane ($\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$, where $\times$ denotes the Cartesian product of sets). We read the above description of $D$ as “the set of all points $(x, y)$ in $\mathbb{R}^2$ such that (or with the property) $x^2 + y^2 \leq 1$.

1.2. Outcomes and the Sample Space. Now we define what we mean by an outcome: An outcome can be a “miss target,” in which case the dart misses the board entirely, or it can be its location in the case that it does not miss the board. Note that we intentionally do not care where the dart lands as long as it missed the board. (The definition of an outcome is totally arbitrary and therefore it is not unique for any experiment.) Mathematically, we form what is called the sample space as the set containing all possible outcomes. If we call this set $\Omega$, then $\Omega = \text{“miss”} \cup D$, where the symbol $\cup$ denotes set union (we say $x \in A \cup B$ if and only if $x \in A$ or $x \in B$). We write $\omega \in \Omega$ to denote a specific outcome from the sample space $\Omega$. For example, $\omega = \text{“miss,”} \omega = (0, 0)$, and $\omega = (0.1, 0.2)$ are possible outcomes; however, according to our definition of an outcome, $\omega$ cannot be $(1, 1)$.

1.3. Events. An event is a collection of outcomes, that is, a subset in $\Omega$. Such a subset can be associated with a question that we may ask about the outcome of the experiment and whose answer can be determined after the outcome is revealed. For example, the question: “Q1: Did the dart land within 0.5 of the bull’s eye?” can be associated with the subset of $\Omega$ (or event) given by $E = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1/4\}$. We would like to call the set $E$ an event. Now consider the complement of Q1, that is: Q2: “Did the dart not land within 0.5 of the bull’s eye?” with which we can associate the event $E^c$, were the superscript “c” denotes set complementation (relative to $\Omega$). Namely, $E^c = \Omega \setminus E$, the set of outcomes (or points or members) of $\Omega$ that are not already in $E$. The notation “\” represents set difference (or subtraction). Note that $E^c = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 > 1/4\} \cup \{\text{“miss”}\}$. The point here is that if $E$ is an event, then we would want $E^c$ to qualify as an event as well since we would like to be able to ask the logical negative of any question. In addition, we would also like to be able to form a logical “or” of any two questions about the experiment outcome. Thus, if $E_1$ and $E_2$ are events, we would like their union to also be an event. For example,
for each $n = 1, 2, \ldots$, define the subset $E_n = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1 - 1/n\}$ and let $\bigcup_{n=1}^{\infty} E_n$ be their countable union. (Notation: $\omega \in E$ if and only if $x \in E_n$ for some $n$.) It is not hard to see (prove it) that $E = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$, which corresponds to the valid question “did the dart land inside the board?” Thus, we would want $E$ (which is the countable union of events) to be an event as well.

Finally, we should be able to ask whether or not the experiment was conducted or not, that is, we would like to label the sample space $\Omega$ as an event as well. With this (hopefully) motivating introduction, we proceed to formally define what we mean by events.

**Definition 1.** A collection $F$ of subsets of $\Omega$ is called a $\sigma$-algebra (read sigma algebra) if:

1. $\Omega \in F$
2. $\bigcup_{n=1}^{\infty} E_n$ whenever $E_n \in F, n = 1, 2, \ldots$
3. $E^c \in F$ whenever $E \in F$

If $F$ is a $\sigma$-algebra, then its members are called events.

Consequences (prove all of them):

1. $\emptyset \in F$ (prove it).
2. $\bigcap_{n=1}^{\infty} E_n \in F$ whenever $E_n \in F, n = 1, 2, \ldots$ (Prove it). Here, the countable intersection $\bigcap_{n=1}^{\infty} E_n$ is defined as follows: $\omega \in \bigcap_{n=1}^{\infty} E_n$ if and only if $x \in E_n$ for all $n$.
3. $A \setminus B \in F$ whenever $A, B \in F$.

Members of a $\sigma$-algebra are also called measurable sets.

**Definition 2.** Let $\Omega$ be a sample space and let $F$ be a $\sigma$-algebra of events. We call the pair $(\Omega, F)$ a measurable space.

**Definition 3.** A collection $\mathcal{D}$ of subsets of $\Omega$ is called a sub-$\sigma$-algebra of $\mathcal{D}$ if

1. $\mathcal{D} \subset F$ (this means that if $A \in \mathcal{D}$, then automatically $A \in F$).
2. $\mathcal{D}$ is itself a $\sigma$-algebra.

**Example 2.** $\{\emptyset, \Omega\}$ is a sub-$\sigma$-algebra of any other $\sigma$-algebra.

**Example 3.** The power set of $\Omega$, which is the set of all subsets of $\Omega$, is a $\sigma$-algebra. In fact it is a maximal $\sigma$-algebra in a sense that it contains any other $\sigma$-algebra. The power set of a set $\Omega$ is often denoted by $2^\Omega$.

**Interpretation:** Once again, we emphasize that it would be meaningful to think of a $\sigma$-algebra as a collection of all valid questions that one may ask about an experiment. The collection has to satisfy certain self-consistency rules, dictated by the requirements for a $\sigma$-algebra, but what we mean by “valid” is really up to us as long as the self-consistency rules are met.
Generation of $\sigma$-algebras: Let $M$ be a collection of events (not necessarily a $\sigma$-algebra). This collection could be a collection of certain events of interest. For example, $\mathcal{M} = \{\{\text{“miss”}\}, \{(x, y) \in \mathbb{R}^2 : 1/4 \leq x^2 + y^2 = 1/2\}\}$ in the dart experiment. The question is can we construct a minimal $\sigma$-algebra that contains $\mathcal{M}$? If such a $\sigma$-algebra exists, call it $\mathcal{F}_M$, then it must possess the following properties:

1. $\mathcal{M} \subseteq \mathcal{F}$
2. If $\mathcal{D}$ is another $\sigma$-algebra containing $\mathcal{M}$, then necessarily $\mathcal{F}_M \subseteq \mathcal{D}$. Hence $\mathcal{F}_M$ is minimal.

The following Theorem states that there is such a minimal $\sigma$-algebra.

**Theorem 1.** Let $\mathcal{M}$ be a collection of events, then there is a minimal $\sigma$-algebra containing $\mathcal{M}$.

Before we prove the Theorem let us look at an example.

**Example 4.** Suppose that $\Omega = (-\infty, \infty)$, and $\mathcal{M} = \{(-\infty, 1), (0, \infty)\}$. It is easy to see (as done in class) that $\mathcal{F}_M = \{\emptyset, \Omega, (-\infty, 1), (0, 1), (-\infty, 0), [1, \infty), (-\infty, 0][1, \infty)\}$. Explain where each member is coming from?

**Proof of Theorem:** Let $\mathcal{R}_M$ be the collection of all $\sigma$-algebras that contain $\mathcal{M}$. We observe that such a collection is not empty since at least it contains the power set $2^\Omega$. Let us label each member of $\mathcal{R}_M$ by an index $\alpha$, namely $D_\alpha$, where $\alpha \in I$, where $I$ is some index set. Define $\mathcal{F}_M = \bigcap_{\alpha \in I} D_\alpha$. We need to show that 1) $\mathcal{F}_M$ is a $\sigma$-algebra containing $\mathcal{M}$, and 2) that $\mathcal{F}_M$ is a minimal $\sigma$-algebra. Note that each $D_\alpha$ contains $\Omega$, thus $\mathcal{F}_M$ contains $\Omega$. If $A \in \mathcal{F}_M$, then $A \in D_\alpha$, for each $\alpha \in I$. Thus, $A^c \in D_\alpha$, for each $\alpha \in I$ (since each $D_\alpha$ is a $\sigma$-algebra) which implies that $A^c \in \bigcap_{\alpha \in I} D_\alpha$. Now suppose that $A_1, A_2, \ldots \in \mathcal{F}_M$. Then, we know that $A_1, A_2, \ldots \in D_\alpha$, for each $\alpha \in I$. Moreover, $\bigcup_{n=1}^{\infty} A_n \in D_\alpha$, for each $\alpha \in I$ (again, since each $D_\alpha$ is a $\sigma$-algebra) and thus $\bigcup_{n=1}^{\infty} A_n \in \bigcap_{\alpha \in I} D_\alpha$. This completes proving that $\mathcal{F}_M$ is a $\sigma$-algebra. Now suppose that $\mathcal{F}'_M$ is another $\sigma$-algebra containing $\mathcal{M}$, we will show that $\mathcal{F}_M \subseteq \mathcal{F}'_M$. First note that since $\mathcal{R}_M$ is the collection of all $\sigma$-algebras containing $\mathcal{M}$, then it must be true that $\mathcal{F}'_M = D_\alpha^\prime$ for some $\alpha' \in I$. If $A \in \mathcal{F}_M$, then necessarily $A \in D_\alpha' = \mathcal{F}'_M$, which establishes that $\mathcal{F}_M \subseteq \mathcal{F}'_M$. This completes the proof of the theorem.

**Example 5.** Let $\mathcal{U}$ be the collection of all open sets in $\mathbb{R}$. Then, according to the above theorem, there exists a minimal $\sigma$-algebra containing $\mathcal{U}$. This $\sigma$-algebra is called the Borel $\sigma$-algebra, $\mathcal{B}$, and its elements are called the Borel subsets of $\mathbb{R}$. Note that by virtue of set complementation, union and intersection, $\mathcal{B}$ contains all closed sets, half open intervals, their countable unions, intersections, and so on.
(Reminder: A subset \( U \) in \( \mathbb{R} \) is called open if for every \( x \in U \), there exists a disc centered at \( x \) which lies entirely in \( U \). Closed sets are defined as complements of open sets. These definitions extend to \( \mathbb{R}^n \) in a straightforward manner.)

**Restrictions of \( \sigma \)-algebras:** Let \( \mathcal{F} \) be a \( \sigma \)-algebra. For any measurable set \( U \), we define \( \mathcal{F} \cap U \) as the collection of all intersections between \( U \) and the members of \( \mathcal{F} \), that is \( \mathcal{F} \cap U = \{ V \cap U : V \in \mathcal{F} \} \). It is easy to show that \( \mathcal{F} \cap U \) is also a \( \sigma \)-algebra, which is called the restriction of \( \mathcal{F} \) to \( U \). (See Exercise 2 in HW#1 and its solution for an example.)

**Example 6.** Back to the dart experiment: What is a reasonable \( \sigma \)-algebra for this experiment? I would say any such \( \sigma \)-algebra should contain all the Borel subsets of \( D \) and the \( \{ \text{miss}\} \) event. So we can take \( \mathcal{M} = \{ \{ \text{miss}\}, \mathcal{B} \cap D \} \). It is easy to check that in this case \( \mathcal{F}_\mathcal{M} = (\mathcal{B} \cap D) \cup \{ \text{miss}\} \cup (\mathcal{B} \cap D) \), where for any \( \sigma \)-algebra \( \mathcal{F} \) and any measurable set \( U \), we define \( \mathcal{F} \cup U \) as the collection of all unions between \( U \) and the members of \( \mathcal{F} \), that is \( \mathcal{F} \cup U = \{ V \cup U : V \in \mathcal{F} \} \). Note that \( \mathcal{F} \cup U \) is not always a \( \sigma \)-algebra (contrary to \( \mathcal{F} \cap U \)), but in this example it is because \( \{ \text{miss}\} \) is the complement of \( D \).

1.4. **Random Variables.** Motivation: Recall the dart experiment, and define the following transformation on the sample space:

\[
X: \Omega \rightarrow \mathbb{R} \text{ defined as }
\]

\[
X(\omega) = \begin{cases} 
1, & \text{if } \omega \in D \\
0, & \text{if } \omega = \{ \text{miss}\}
\end{cases}
\]

where \( D = \{ (x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1 \} \). Consider the sets of outcomes that we can identify if we knew that \( X \) fell in the interval \((\infty, r)\). More precisely, we want to identify \( \{ \omega \in \Omega : X(\omega) \in (\infty, r) \} \), or equivalently, the set \( X^{-1}(\infty, r) \), which is the inverse image of the set \((\infty, r)\). It is easy to see that:

For \( r < 0 \), \( X^{-1}(\infty, r)) = \emptyset \),

for \( 0 \leq r < 1 \), \( X^{-1}(\infty, r)) = \{ \text{miss}\} \)

and for \( 1 \leq r < \infty \), \( X^{-1}(\infty, r)) = \Omega \).

Note that in each case, \( X^{-1}(\infty, r)) \) is an event (i.e., a member of \( \mathcal{F} \) for this experiment). Now let \( \mathcal{M} = \{ \emptyset, \{ \text{miss}\}, \Omega \} \), then it is easy to check that \( \mathcal{F}_\mathcal{M} = \{ \emptyset, \{ \text{miss}\}, \Omega, D \} \), which can be intuitively identified as the set of all events that \( X \) can convey about the experiment. In particular, \( \mathcal{F}_\mathcal{M} \) consists of precisely those events whose occurrence can be determined through our observation of the value of \( X \). In other words, \( \mathcal{F}_\mathcal{M} \) is the “information” that \( X \) can provide about the outcome of the experiments. Note that \( \{ \emptyset, \{ \text{miss}\}, \Omega, D \} \) is much smaller than the original \( \sigma \)-algebra \( \mathcal{F} \), which was \( (\mathcal{B} \cap D) \cup \{ \text{miss}\} \cup (\mathcal{B} \cap D) \). Clearly, \( \{ \emptyset, \{ \text{miss}\}, \Omega, D \} \subset (\mathcal{B} \cap D) \cup \{ \text{miss}\} \cup (\mathcal{B} \cap D) \).
Thus, $X$ only partially informs us of the true outcome of the experiment; albeit, whatever $X$ conveys about the experiment is an event.

Motivated by this example, we proceed to define what we mean by a random variable.

**Definition 4.** Let $(\Omega, \mathcal{F})$ be a measurable space. A transformation $X : \Omega \to \mathbb{R}$ is said to be $\mathcal{F}$-measurable if for every $r \in \mathbb{R}, X^{-1}((-\infty, r)) \in \mathcal{F}$. In such a case, $X$ is called a random variable.

Now let $X$ be a random variable and consider the collection of events $\mathcal{M} = \{\omega \in \Omega : X(\omega) \in (-\infty, r), r \in \mathbb{R}\}$, which can also be written as $\{X^{-1}((-\infty, r)), r \in \mathbb{R}\}$. As before, let $\mathcal{F}_\mathcal{M}$ be the minimal $\sigma$-algebra containing $\mathcal{M}$. Then, $\mathcal{F}_\mathcal{M}$ is the “information” that the random variable $X$ conveys about the experiment. We define such a $\sigma$-algebra from this point on as $\sigma(X)$, the $\sigma$-algebra generated by the random variable $X$. In the above example of $X$, $\sigma(X) = \{\emptyset, \{\text{miss}\}, \Omega, D\}$. See Exercise 3 in HW#1 for another example of $\sigma(X)$.

**Facts about Measurable Transformations:** Let $(\Omega, \mathcal{F})$ be a measurable space. The following statements are equivalent:

1. $X$ is a measurable transformation.
2. $X^{-1}((-\infty, r]) \in \mathcal{F}$. (we proved this in class)
3. $X^{-1}((r, \infty)) \in \mathcal{F}$.
4. $X^{-1}([r, \infty)) \in \mathcal{F}$.
5. $X^{-1}((a, b)) \in \mathcal{F}$ for all $a \leq b$.
6. $X^{-1}((a, b]) \in \mathcal{F}$ for all $a \leq b$.
7. $X^{-1}([a, b]) \in \mathcal{F}$ for all $a \leq b$.
8. $X^{-1}((a, b)) \in \mathcal{F}$ for all $a \leq b$.
9. $X^{-1}(B) \in \mathcal{F}$ for all $B \in \mathcal{B}$.

Using (9), we can equivalently define $\sigma(X)$ as $\{X^{-1}(B) : B \subset \mathcal{B}\}$, which can be directly shown to be a $\sigma$-algebra (prove it).
2.1. Probability Measure.

**Definition 5.** Consider the measurable space \((\Omega, \mathcal{F})\). A set function, \(P\), mapping \(\mathcal{F}\) into \(\mathbb{R}\) is called a *probability measure* if

1. \(P(E) \geq 0\).
2. \(P(\Omega) = 1\).
3. If \(E_1, E_2, \ldots \in \mathcal{F}\) and if \(E_i \cap E_j = \emptyset\) when \(i \neq j\), then \(P(\bigcup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} P(E_n)\).

The following properties follow directly from the above definition.

**Property 1.** \(P(\emptyset) = 0\).

**Proof.** Put \(E_1 = \Omega, E_2 = \ldots = E_n = \emptyset\) in (3) and use (2) to get \(1 = P(\Omega) = P(\Omega \cup \emptyset \cup \emptyset \cup \ldots) = P(\Omega) + \sum_{n=2}^{\infty} P(\emptyset) = 1 + \sum_{n=2}^{\infty} P(\emptyset)\). Thus, \(\sum_{n=2}^{\infty} P(\emptyset) = 0\), which implies that \(P(\emptyset) = 0\) since \(P(\emptyset) \geq 0\) according to (1).

**Property 2.** If \(E_1, E_2, \ldots, E_n \in \mathcal{F}\) and if \(E_i \cap E_j = \emptyset\) when \(i \neq j\), then \(P(\bigcup_{i=1}^{n} E_i) = \sum_{i=1}^{n} P(E_i)\).

**Proof.** Put \(E_{n+1} = E_{n+2} = \ldots = \emptyset\) and the result will follow from 3 since \(P(\emptyset) = 0\) (from Property 2).

**Property 3.** If \(E_1, E_2 \in \mathcal{F}\) and \(E_1 \subseteq E_2\), then \(P(E_1) \leq P(E_2)\).

**Proof.** Note that \(E_1 \cup E_2 \setminus E_1 = E_2\) and \(E_1 \cap E_2 \setminus E_1 = \emptyset\). Thus, by Property 2 (use \(n = 2\)), \(P(E_2) = P(E_1) + P(E_2 \setminus E_1) \geq P(E_1)\), since \(P(E_2 \setminus E_1) \geq 0\).

**Property 4.** If \(A_1 \subseteq A_2 \subseteq A_3 \ldots \in \mathcal{F}\), then

\[
\lim_{n \to \infty} P(A_n) = P\left(\bigcup_{n=1}^{\infty} A_n\right).
\]

**Proof.** Put \(B_1 = A_1, B_2 = A_2 \setminus A_1, \ldots, B_n = A_{n+1} \setminus A_n, \ldots\). Then, it is easy to check that \(\bigcup_{n=1}^{\infty} B_n\) and \(\bigcup_{n=1}^{m} B_n = \bigcup_{n=1}^{m} A_n \setminus \bigcup_{m+1}^{\infty} A_n\). Then, \(B_i \cap B_j = \emptyset\) when \(i \neq j\). Hence, \(P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(B_i)\) and \(P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(B_i)\). But, \(\bigcup_{i=1}^{n} A_i = A_n\), so that \(P(A_n) = \sum_{i=1}^{n} P(B_i)\). Now since \(\sum_{i=1}^{n} P(B_i)\) converges to \(\sum_{i=1}^{\infty} P(B_i)\), we conclude that \(P(A_n)\) converges to \(\sum_{i=1}^{\infty} P(B_i)\), which is equal to \(P(\bigcup_{i=1}^{\infty} A_i)\) as shown above.

**Property 5.** If \(A_1 \supset A_2 \supset A_3 \ldots\), then \(\lim_{n \to \infty} P(A_n) = P(\bigcap_{n=1}^{\infty} A_n)\).

**Proof.** See HW#2.

The triplet \((\Omega, \mathcal{F}, P)\) is called a *probability space*. 
Example 7. Recall the dart experiment. We now define \( P \) on \((\Omega, F)\). Assign \( P(\{ \text{"miss"} \}) = 0.5 \), and for \( A \in D \cap \mathfrak{B} \), assign \( P(A) = \text{area}(A)/2\pi \). It is easy to check that \( P \) defines a probability measure. (For example, \( P(\Omega) = P(D \cup \{ \text{"miss"} \}) = P(D) + P(\{ \text{"miss"} \}) = \text{area}(D)/2\pi + 0.5 = 0.5 + 0.5 = 1 \). Check the other requirements as an exercise.)

Distributions: Consider a probability space \((\Omega, F, P)\), and consider a random variable \( X \) defined on it. Up to this point, we have developed a formalism that allows us to ask questions of the form “what is the probability that \( X \) assumes a value in a Borel set \( B \)?” Symbolically, this is written as \( P(\{ \omega \in \Omega : X(\omega) \in B \}) \), or for short, \( P \{ X \in B \} \) with the understanding that the set \( \{ X \in B \} \) is an event (i.e., member of \( F \)). Answering all the questions of the above form is tantamount to assigning a number in the interval \([0,1]\) to every Borel. Thus, we can think of a mapping from \( B \) into \([0,1]\) that provides an answer to all the questions of the form described earlier. We call this mapping the distribution of the random variable \( X \), and it is denoted by \( X \). Formally, we have \( \mu_X : \mathfrak{B} \to [0,1] \) according to the rule \( \mu_X(B) = P \{ X \in B \}, B \in \mathfrak{B} \).

Proposition 1. \( \mu_X \) defines a measure on \( \mathfrak{B} \).

Proof. See HW#4.

Distribution Functions: Recall from your undergraduate probability that we often associate with each random variable a distribution function, defined as \( F_X(x) = P \{ X \leq x \} \). This function can also be obtained from the distribution of \( X \), \( \mu_X \), by evaluating \( \mu_X \) at \( B = (-\infty, x] \), which is a Borel set. That is, \( F_X(x) = \mu_X((-\infty, x]) \).

Property 6. \( F_X \) is nondecreasing.

Proof. For \( x_1 \leq x_2, (-\infty, x_1] \subseteq (-\infty, x_2] \) and \( = \mu_X((-\infty, x_1]) \leq \mu_X((-\infty, x_2]) \) since \( \mu_X \) is a probability measure (see Property 3 above).

Property 7. \( \lim_{x \to \infty} F_X(x) = 1 \) and \( \lim_{x \to -\infty} F_X(x) = 0 \).

Proof. Note that \((-\infty, \infty) = \bigcup_{n=1}^{\infty} (-\infty, n] \) and by Property 4 above, \( \lim_{n \to \infty} \mu_X((-\infty, n]) = \mu_X((-\infty, \infty)) = 1 \) since \( \mu_X \) is a probability measure. Thus we proved that \( \lim_{n \to \infty} F_X(n) = 1 \). Now the same argument can be repeated if we replace the sequence \( n \) by any increasing sequence \( x_n \). Thus, \( \lim_{x \to -\infty} F_X(x_n) = 1 \) for any increasing sequence \( x_n \), and consequently \( \lim_{x \to \infty} F_X(x) = 1 \).

The proof of the second assertion is left as an exercise.

Property 8. \( F_X \) is right continuous, that is, \( \lim_{x \uparrow y} F_X(x) = F_X(y) \).
Proof. Note that \( F_X(y) = \mu_X((-\infty, y]) \), and \((-\infty, y] = \bigcup_{n=1}^{\infty} (-\infty, y + n^{-1}] \). So, by Property 4 above, \( \lim_{n \to \infty} \mu_X((-\infty, y + n^{-1}]) = \mu_X((-\infty, y + n^{-1}]) = \mu_X((-\infty, y]]. \) Thus, we proved that \( \lim_{n \to \infty} F_X(y + n^{-1}) = F_X(y) \). In the same fashion, we can generalize the result to obtain \( \lim_{n \to \infty} F_X(y + x_n) = F_X(y) \) for any sequence for which \( x_n \downarrow 0 \). This completes the proof.

**Example 8.** (Problem 3.8 in the textbook) Let \( X \) be a uniformly-distributed random variable in \([0, 2]\). Let the function \( g \) be as shown in Fig. 1 below. Compute the distribution function of \( Y = g(X) \).

![Graph of g(x) and F(y)](image)

**Solution:** If \( y < 0 \), \( F_Y(y) = 0 \) since \( Y \) is nonnegative. If \( 0 \leq y \leq 1 \), \( F_Y(y) = P\{X \leq y/2\} \cup \{X > 1 - y/2\} = 0.5y + 0.5 \). Finally, if \( y > 1 \), \( F_Y(y) = 1 \) since \( Y \leq 1 \). The graph of \( F_Y(y) \) is shown above. Note that \( F_Y(y) \) is indeed right continuous.

### 2.2. Expectation

Recall that in an undergraduate probability course one would talk about the expectation, average, or mean of a random variable. This is done by carrying out an integration (in the Riemann sense) with respect to a probability density function. It turns out that the definition of an expectation does require having a probability density function (pdf). It is based on a more or less intuitive notion of an average. We will follow this general approach here and then connect it to the usual expectation with respect to a pdf whenever the pdf exists. We begin by introducing the expectation of a nonnegative random variable, and will generalize thereafter.

Consider a nonnegative random variable \( X \), and for each \( n \geq 1 \), we define the sum \( S_n = \sum_{i=1}^{\infty} \frac{i}{2^n} P\{\frac{i}{2^n} < X \leq \frac{i+1}{2^n}\} \). We claim that \( S_n \) is nondecreasing. If this is the case (to be proven shortly), then we know that \( S_n \) is either convergent (to a finite number) or \( S_n \uparrow \infty \). In any case, we call the limit of \( S_n \) the expectation of \( X \), and symbolically we denote as \( E[X] \). Thus, \( E[X] = \lim_{n \to \infty} S_n \). To see the monotonicity of \( S_n \), we follow Chow and Teicher [1] and observe that

\[
\{\frac{i}{2^n} < X \leq \frac{i+1}{2^n}\} = \{\frac{2i}{2^n} < X \leq \frac{2i+2}{2^n}\} = \{\frac{2i}{2^n} < X \leq \frac{2i+1}{2^n}\} \cup \{\frac{2i+1}{2^n} < X \leq \frac{2i+2}{2^n}\};
\]

thus,

\[
S_n = \sum_{i=1}^{\infty} \frac{2i}{2^n} P\{\frac{2i}{2^n} < X \leq \frac{2i+1}{2^n}\} + P\{\frac{2i+1}{2^n} < X \leq \frac{2i+2}{2^n}\};
\]

thus,

\[
\leq \frac{1}{2^n} P\{\frac{1}{2^n} < X \leq \frac{2}{2^n}\} + \sum_{i=1}^{\infty} \frac{2i}{2^{n+1}} P\{\frac{2i}{2^n} < X \leq \frac{2i+1}{2^n}\} + P\{\frac{2i+1}{2^n} < X \leq \frac{2i+2}{2^n}\}\]

Therefore, the limit is defined as

\[
E[X] = \lim_{n \to \infty} \sum_{i=1}^{\infty} \frac{2i}{2^n} P\{\frac{2i}{2^n} < X \leq \frac{2i+1}{2^n}\}
\]

Thus, \( E[X] \) is defined as the limit of \( S_n \) and is called the expectation of \( X \).
that if $E$ is a nonnegative random variable to define $E[X] = E[X^+] \cdot E[X^-]$ whenever $E[X^+] < \infty$ or $E[X^-] < \infty$, or both. In cases where $E[X^+] < \infty$ and $E[X^-] = \infty$, or $E[X^-] < \infty$ and $E[X^+] = \infty$, we define $E[X] = -\infty$ and $E[X] = \infty$, respectively. Finally, $E[X]$ is not defined whenever $E[X^+] = E[X^-] = \infty$.

Special Case: Suppose that $X$ is a binary random variable, that is, $X(\omega) = 1$ if $\omega \in E$ and $I_E(\text{omega}) = 0$ otherwise, where $E$ is some event. Then, $E[X] = P(E)$. (Prove it.)

Notation and Terminology: $E[X]$ is also written as $\int_X X(\omega)P(d\omega)$, which is called the Lebesgue integral of $X$ with respect to the probability measure $P$. Often, cumbersome notation is avoided by writing $\int_X X P(d\omega)$ or simply $\int_X X dP$.

Linearity of Expectation: The expectation is linear, that is, $E[aX+bY] = aE[X]+bE[Y]$. This can be seen, for example, by observing that any nonnegative random variable can be approximated from below by functions of the form $\sum_{i=1}^{\infty} x_i I_{\{x_i < X \leq x_{i+1}\}}(\omega)$, where for any event $E$, the random variable $I_E(\omega) = 1$ if $\omega \in E$ and $I_E(\omega) = 0$ otherwise. ($I_E$ is called the indicator function of the set $E$.) Indeed, we have seen such an approximation through our definition of the expectation. Namely, if we define $X_n(\omega) = \sum_{i=1}^{\infty} \frac{1}{2^n} I_{\{\frac{i}{2^n} < X \leq \frac{i+1}{2^n}\}}(\omega)$, then it is easy to check that $X_n(\omega) \to X(\omega)$, as $n \to \infty$. In fact, $E[X]$ was defined as $\lim_{n \to \infty} E[X_n]$, where $E[X_n]$ precisely coincides with $S_n$ described above [recall that $E[I_X(\omega)] = P(E)$]. Now to prove the linearity of expectations, we note that if $X$ and $Y$ are random variables with defined expectations, then we can approximate them by $X_n$ and $Y_n$, respectively. Also, $X_n + Y_n$ would approximate $X + Y$. Next, we observe that

$$E[X_n + Y_n] = E[\sum_{i=1}^{\infty} \frac{1}{2^n} I_{\{\frac{i}{2^n} < X \leq \frac{i+1}{2^n}\}}(\omega)] + E[\sum_{i=1}^{\infty} \frac{1}{2^n} I_{\{\frac{i}{2^n} < Y \leq \frac{i+1}{2^n}\}}(\omega)]$$

$$= E[\sum_{i=1}^{\infty} \frac{1}{2^n} I_{\{\frac{i}{2^n} < X \leq \frac{i+1}{2^n}\}}(\omega) \cap \{\frac{i}{2^n} < Y \leq \frac{i+1}{2^n}\}] + E[\sum_{i=1}^{\infty} \frac{1}{2^n} I_{\{\frac{i}{2^n} < X \leq \frac{i+1}{2^n}\}}(\omega) \cap \{\frac{i}{2^n} < Y \leq \frac{i+1}{2^n}\}]$$

$$= E[\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \frac{1}{2^n} I_{\{\frac{i+1}{2^n} < \frac{j}{2^n} \leq \frac{i+2}{2^n}\}}(\omega)]$$

But $\sum_{j=1}^{\infty} \sum_{i=1}^{\infty} \frac{1}{2^n} I_{\{\frac{i}{2^n} < X \leq \frac{i+1}{2^n}\}}(\omega) \cap \{\frac{i}{2^n} < Y \leq \frac{i+1}{2^n}\}(\omega) = E[X_n + Y_n]$. 

If $E[X] < \infty$, we say that $X$ is integrable.
Thus, we have shown that $E[X_n] + E[Y_n] = E[X_n + Y_n]$. Now take limits of both sides and use the definition of $E[X]$, $E[Y]$ and $E[X + Y]$ as the limits of $E[X_n]$, $E[Y_n]$, and $E[X_n + Y_n]$, respectively, to conclude that $E[X] + E[Y] = E[X + Y]$. The homogeneity property $E[aX] = aE[X]$ can be proved similarly.

**Expectations in the Context of Distributions:** Recall that for a nonnegative random variable $X$, $E[X] = \lim_{n \to \infty} \sum_{i=1}^{\infty} \frac{i}{2^n} P\{\frac{i}{2^n} < X \leq \frac{i+1}{2^n}\}$. But we had seen earlier that $P\{\frac{i}{2^n} < X \leq \frac{i+1}{2^n}\} = \mu_X((\frac{i}{2^n}, \frac{i+1}{2^n}))$. So we can write $\sum_{i=1}^{\infty} \frac{i}{2^n} P\{\frac{i}{2^n} < X \leq \frac{i+1}{2^n}\}$ as $\sum_{i=1}^{\infty} \frac{i}{2^n} \mu_X((\frac{i}{2^n}, \frac{i+1}{2^n}))$. We denote the limit of the latter by $\int_{\Omega} x d\mu_X$, which is read as the *Lebesgue integral of $x$ with respect to the probability measure $\mu_X$*. In summary, we have $E[X] = \int_{\Omega} X dP = \int_{\Omega} x d\mu_X$. 
3. Lecture 1: 08/25/03

3.1. Definition of Sample Space, Events and Random Variables.

Example 9. Shoot a dart at a target $D$, as depicted in Fig. 2. $D = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$.

Outcome:

(1) “miss”

(2) Give coordinates on the board $D$.

Figure 2.

The collection of all outcomes (if it is a set) is called a sample space associated with the experiment. We call this set $\Omega$ and designate its members generically by $\omega$ ($\omega \in \Omega$), e.g., “miss” $\in \Omega$, $(0.5, \pi/4) \in \Omega$, and $(0, 0) \in \Omega$. We can group outcomes and form subsets of $\Omega$, e.g., $E_1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1/2\} \subset \Omega$. Subsets of $\Omega$ are “called” events. We can think of $E_1$ as a question: $Q_1 = “Is$ the point within $\sqrt{1/2}$ feet or less from the bullseye?” $Q_2 = “Did I miss?”$ $E_2 = \{ “miss” \}$. Clearly $E_2 \subset \Omega$ (subset of) because $\Omega = \{ “miss” \} \cup \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$. $Q_3$: $Q_1$ or $Q_2$? $E_3 = E_1 \cup E_2$.

Remark: If $E_1$ and $E_2$ are events then we want their union to be an event as well. $Q_1^c = “Did$ we hit outside the disc of radius $\sqrt{1/2}?”$, where the “c” means set complementation.

Remark: If $E$ is an event then we would like the complement $E^c$ to be an event as well.

Consequence: If $E_1$ and $E_2$ are events, is $E_1 \cap E_2$ an event? Yes. Because:

(1) $E_1 \cap E_2 = \emptyset = \{\}$

$E_1 \cap E_2$ is the impossible event (never occurs).

(2) $E_1 \cap E_2 = (E_1^c \cup E_2^c)^c$

Let’s think of a legitimate collection of events. Call it $\mathcal{F}$, e.g., $\mathcal{F} = \{E_1, E_2, E_1 \cup E_2, \emptyset, E_1^c, E_2^c, \Omega, (E_1 \cup E_2)^c\}$. $\mathcal{F}$ must have the following properties:

(1) $\emptyset \in \mathcal{F}$.

(2) If $A, B \in \mathcal{F}$ then $A \cup B \in \mathcal{F}$.

(3) If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$.

We are not done: Suppose we take $n = 1, 2, 3, \ldots$. $A_n = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1/2 - 1/n\}$. Note that $A_n \subset A_{n+1}$. If $A_n$ occurs then $A_{n+1}$ occurs.
Define: \( A_\infty = \bigcup_{n=1}^{\infty} A_n \). By this we mean: if \( \omega \in A \) then \( \omega \in A_i \) for some \( i \) and if \( \omega \in A_i \) for some \( i \), then \( \omega \in A \). \( A_\infty = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1/2\} \), which is a logically legit event.

**Formal Definition:** Suppose that \( \mathcal{F} \) is a collection of subsets of \( \Omega \), then we call \( \mathcal{F} \) a \( \sigma \)-algebra (of sets) if:

1. \( \Omega \in \mathcal{F} \).
2. If \( A \in \mathcal{F} \), then \( A^c \in \mathcal{F} \).
3. If \( A_1, A_2, \ldots \in \mathcal{F} \), then \( \bigcup_{n=1}^{\infty} A_n \in \mathcal{F} \).

### 3.2. Sub-algebras.

We call a collection of events \( \mathcal{D} \) a sub-algebra of \( \mathcal{F} \) if \( \mathcal{D} \) is a \( \sigma \)-algebra and \( \mathcal{D} \subseteq \mathcal{F} \).

**Generation of \( \sigma \)-algebras:** Let \( \mathcal{M} \) be a collection of events [(not necessarily a \( \sigma \)-algebra) for example \( \mathcal{M} = \{E_1, E_2\} \)], then there is a minimal \( \sigma \)-algebra \( \mathcal{F}_\mathcal{M} \) that contains \( \mathcal{M} \).

**Sketch of Proof:** We will be done if we have a \( \sigma \)-algebra that contains \( \mathcal{M} \) and any other such \( \sigma \)-algebra is necessarily contained in it.

1. Let \( \tilde{K} \) be the collection of all \( \sigma \)-algebras that contain \( \mathcal{M} \). \( \tilde{K} \) is not empty because the power set (the set of all subsets) is a \( \sigma \)-algebra.
2. Take \( \mathcal{F}_\mathcal{M} = \) the intersection of all members of \( \tilde{K} \) and show that it is a \( \sigma \)-algebra.
4. Lecture 2: 08/27/03

4.1. Set Theory Notation.

(1) If $A, B$ are sets then $A - B = \{ x \in A : x \notin B \}$.

(2) $2^A$ is the collection of all subsets of $A$ (including $\emptyset$ and the set itself).

(3) If $A_1, A_2, \ldots$ are sets, then $A = \bigcup_{n=1}^{\infty} A_n$ means $x \in A$ iff $x \in A_i$ for some $i$.

(4) $B = \bigcap_{n=1}^{\infty} A_n$ means

\[
\begin{array}{c}
A \\
\cap
\end{array}
\]

\[
\begin{array}{c}
\cup
\end{array}
\]

Theorem 2. There is a minimal $\sigma$-algebra $\mathcal{F}_M$ that contains $\mathcal{M}$.

Proof. Let $\mathcal{R}_M$ be the collection of all $\sigma$-algebras that contain $\mathcal{M}$. (e.g., $2^M$ is one such $\sigma$-algebra.) Let $I$ be an index set for $\mathcal{R}_M$. $\mathcal{R}_M = \{ \mathcal{D}_\alpha, \alpha \in I \}$, where $\mathcal{D}_\alpha$ is one such $\sigma$-algebra. Put $\mathcal{F}_M = \bigcap_{\alpha \in I} \mathcal{D}_\alpha$.

(1) We show that $\mathcal{F}_M$ is a $\sigma$-algebra.

(a) Does it contain $\emptyset$? Yes, each $\mathcal{D}_\alpha$ contains $\emptyset$ and hence the $\bigcap_{\alpha \in I} \mathcal{D}_\alpha$.

(b) Complementation? If $A \in \mathcal{F}_M$, is $A^c \in \mathcal{F}_M$? Yes, because if $A \in \mathcal{F}_M$, then $A \in \mathcal{D}_\alpha, \forall \alpha \in I$. ($\mathcal{F}_M$ is $\cap$ of all $\mathcal{D}_\alpha$.) Then, $A^c \in \mathcal{D}_\alpha, \alpha \in I$, then $\bigcap_{\alpha \in I} \mathcal{D}_\alpha$ contains $A^c$.

(c) Union? If $A_1, \ldots, A_n, \ldots \in \mathcal{F}_M$, does $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}_M$? If $A_1, \ldots, A_n \in \mathcal{F}_M \Rightarrow A_1, \ldots, A_n \in \bigcap_{\alpha \in I} \mathcal{D}_\alpha \Rightarrow A_1, \ldots, A_n \in \mathcal{D}_\alpha, \forall \alpha$. Since $\mathcal{D}_\alpha$ is a $\sigma$-algebra $\Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{D}_\alpha, \forall \alpha \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \bigcap_{\alpha \in I} \mathcal{D}_\alpha \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}_M$.

Conclusion: $\mathcal{F}_M$ is a $\sigma$-algebra. We show that $\mathcal{F}_M$ is minimal. Suppose that $\mathcal{F}'_M$ is a $\sigma$-algebra containing $\mathcal{M}$. We need to show that $\mathcal{F}_M \subset \mathcal{F}'_M$. Pick one event $A$ in $\mathcal{F}_M$. Then, $A \in \bigcap_{\alpha \in I} \mathcal{D}_\alpha$, then $A \in \mathcal{D}_\alpha \forall \alpha \in I$. But, $\mathcal{F}'_M = \mathcal{D}_\alpha^*$ for some $\alpha^* \in I$. \vdots $A \in \mathcal{D}_\alpha^*, \ldots, A \in \mathcal{F}'_M$. 

Example 10. $A_n = (-1/n, 1/n)$ interval. Then $\bigcap_{n=1}^{\infty} A_n = \{ 0 \}$. Then $\bigcap_{n=1}^{\infty} A_n = \{ -1, 1 \}$.

Sample space: $\Omega$

Power set: $2^\Omega$ $\rightarrow$ trivial $\sigma$-algebra. The power set is the largest $\sigma$-algebra. $\{ \emptyset, \Omega \}$ is the smallest $\sigma$-algebra.

If $\mathcal{F}$ is a $\sigma$-algebra, then members of it are called events or measurable sets.

Let $\mathcal{M}$ be a collection of events.
Example 11. \( \Omega = (-\infty, \infty) \),
\[ \mathcal{F}_M = \{\emptyset, (-\infty, \infty), (-\infty, 1) \cup (-1, \infty) \}. \]

Remark: If \( A, B \in \mathcal{F} \) then \( A - B \) and \( B - A \in \mathcal{F} \) (idea: \( A - B = A \cap B^c \)).

Example 12. \( \mathcal{U} \) is the collection of all open sets in \( \mathbb{R} \). [Reminder: \( \mathcal{U} \) is open if for every \( x \in \mathcal{U} \), we can find \( \varepsilon > 0 \), so that \( (x - \varepsilon, x + \varepsilon) \in \mathcal{U} \).] The theorem tells us that there is a minimal \( \sigma \)-algebra that contains \( \mathcal{U} \). This is called a Borel \( \sigma \)-algebra, and includes all the open sets. \( \mathcal{B} \) contains all open and closed sets (because it contains complements). Members of \( \mathcal{B} \) are called Borel sets.

If \( \mathcal{U} \) is an open set, then \( \mathcal{B} \cap \mathcal{U} \). This means that the collection of all Borel sets each one intersected with \( \mathcal{U} \). This is called the restriction of \( \mathcal{B} \) to \( \mathcal{U} \).

Exercise 1. \( \mathcal{B} \cap \mathcal{U} \) is a \( \sigma \)-algebra.

4.2. Random Variables. Suppose that \( \Omega \), we also have \( \mathcal{F} \). We can define a function \( X \) on \( \Omega \) (really, we have \( X(\omega), \omega \in \Omega \) taking values in \( \mathbb{R} \). Recall the dart example:

Let \( D = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\} \). \( \Omega = \{\text{"miss"}\} \cup D \). Let \( \mathcal{F} \) be the \( \sigma \)-algebra. Define the function \( X(\omega) \) by the rule:

\[
X(\omega) = \begin{cases} 
10, & \text{if } \omega = \text{"miss"} \\ 
\sqrt{x^2 + y^2}, & \text{if } \omega \in D 
\end{cases}
\]

Let’s consider the collection of outcomes that correspond to \( x < 1/2 \). We write the above event as \( \{\omega \in \Omega : X(\omega) < 1/2\} = A_1 = X^{-1}((-\infty, 1/2)) \), which is the inverse image in \( \Omega \) of the outcome.

\[
\begin{align*}
A_1 &= \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1/4\} \in \mathcal{F} \\
A_2 &= X^{-1}((-\infty, 2)) = D \in \mathcal{F} \\
A_3 &= X^{-1}((-\infty, 11)) = D \cup \{\text{"miss"}\} = \Omega \in \mathcal{F} \\
A_0 &= X^{-1}((-\infty, 0]) = \emptyset \in \mathcal{F}
\end{align*}
\]

Formal Definition of a Random Variable: Given \( \Omega, \mathcal{F} \), a mapping \( x : \Omega \to \mathbb{R} \) is called a random variable (r.v.) if \( X^{-1}((-\infty, r)) \in \mathcal{F}, \forall r \in \mathbb{R} \).

Names:

(1) \( (\Omega, \mathcal{F}) \) is called a measurable space.
(2) Such a mapping is also called a $\mathcal{F}$-measurable transformation.

Back to dart example:

(3) $X_1(\omega) = \begin{cases} 
1, & \text{if } \omega \in \mathcal{D} \\
0, & \text{if } \omega = \text{“miss”}
\end{cases}$

(4) $X_2(\omega) = \begin{cases} 
2, & \text{if } \omega \in \{(x, y) : x^2 + y^2 \leq 1/4\} \\
1, & \text{if } \omega \in \{(x, y) : 1/4 < x^2 + y^2 \leq 1\} \\
0, & \text{if } \omega = \text{“miss”}
\end{cases}$

![Figure 6.](image)

Let's form a $\sigma$-algebra from the collection $\emptyset, \{\text{“miss”}\}, \Omega \equiv \mathcal{M}_1$. $\mathcal{F}_{\mathcal{M}_1} = \{\emptyset, \{\text{“miss”}\}, \Omega, \mathcal{D}\}$. $\mathcal{F}_{\mathcal{M}_1}$ conveys all the information that $x$ conveys about the experiment. We write $\sigma(x)$ in place of $\mathcal{F}_{\mathcal{M}_1}$ to emphasize the $\sigma$-algebra (or information) that is generated by the r.v. $x$.

**Exercise 2.** Find $\sigma(X_2)$. 
5. Lecture 3: 09/03/2003

5.1. Random Variables. \( x : \Omega \to \mathbb{R} \) is called a r.v. if \( x \) is \( \mathcal{F} \)-measurable. Meaning: \( X^{-1}((-\infty, r)) \in \mathcal{F}, \forall r \). \( X^{-1}((-\infty, r)) \) is the inverse image/transformation of \((-\infty, r)\) under \( x \).

The \( \sigma \)-algebra generated by the r.v. \( x \) is \( \sigma(x) \) = the smallest \( \sigma \)-algebra that contains all the inverse images of the form \( X^{-1}((-\infty, r)) \), \( \forall r \). Figure 7.

\[ X^{-1}(-\infty, r) \]

\[ (-\infty, r) \]

Let \( \mathcal{M}_X = \{ X^{-1}((-\infty, r)), r \in \mathbb{R} \} \). Because \( X \) is a r.v., for each \( r \), \( X^{-1}((-\infty, r)) \in \mathcal{F} \), \( \mathcal{M}_X \) is a collection of events.

\( \therefore \) By the theorem, there exists a \( \sigma \)-algebra \( \mathcal{F}_{\mathcal{M}_X} \) that minimally contains \( \mathcal{M}_X \) (we call it \( \sigma(x) \)). In general, \( \sigma(x) \subset \mathcal{F} \), but not necessarily equal to \( \mathcal{F} \). Thus, by observing the value of the r.v., we may not be able to answer all the questions embedded in \( \mathcal{F} \). (Think of a “miss” or “hit” type r.v. for the dart example.)

5.2. Properties of Random Variables. The following statements are equivalent:

(1) \( x \) is a r.v.
(2) \( X^{-1}((-\infty, r]) \in \mathcal{F}, \forall r \).
(3) \( X^{-1}((r, \infty)) \in \mathcal{F}, \forall r \).
(4) \( X^{-1}([r, \infty)) \in \mathcal{F}, \forall r \).
(5) \( X^{-1}((a, b)) \in \mathcal{F}, \forall a, b. \quad [(a, b), (a, b], [a, b), [a, b]\]
(6) \( X^B \in \mathcal{F}, \forall B \in \mathcal{B} \).

Property 6 is the most powerful because \((-\infty, r), (-\infty, r], (r, \infty), [r, \infty), (a, b), (a, b], [a, b), [a, b]\) are Borel sets.
5.3. Measure. A set function \( P : \mathcal{F} \to [0,1] \) is called a probability measure (or just probability) if:

(1) \( P(\Omega) = 1 \).

(2) If \( A_1, A_2, A_3 \) are mutually disjoint (meaning \( A_i \cap A_j = \emptyset \) if \( i \neq j \)), then \( P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n) \).

Properties of \( P \):

(1) \( P(\emptyset) = 0; \Omega = \Omega \cup \emptyset \). Note: \( \Omega \) and \( \emptyset \) are disjoint.

\[
P(\Omega) = P(\Omega \cup \emptyset) \\
1 = P(\Omega) + P(\emptyset) = 1 + P(\emptyset) \\
\Rightarrow P(\emptyset) = 0.
\]

(2) If \( A \cup B \in \mathcal{F} \), then \( P(B) \geq P(A) \).

\[
B = A \cup B \setminus A \text{ (disjoint).} \\
P(B) = P(A \cup B \setminus A) = P(A) + P(B \setminus A) \geq P(A)
\]

![Figure 8](image)

(3) If \( A_1 \subset A_2 \subset A_3 \subset A_4 \ldots \), then \( P(\bigcup_{n=1}^{\infty} A_n) = \lim_{n \to \infty} P(A_n) \). \( f(x) \) is continuous at \( x_0 \) if

\[
\lim_{n \to \infty} f(x_n) = f(x_0) \text{ if } x_n \to x_0 \text{ as } n \to \infty.
\]

![Figure 9](image)

Proof: Put \( B_1 = A_1, B_2 = A_2 \setminus A_1, \ldots, B_n = A_n \setminus A_{n-1} \). Now, \( \bigcup_{i=1}^{n} B_i = \bigcup_{i=1}^{n} A_i = A_n \). Also,

\[
\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i.
\]

\( B_i \)'s are disjoint.

\[
\therefore P(\bigcup_{i=1}^{\infty} P(B_i) = \sum_{i=1}^{n} P(B_i) = P(A_n), \text{ true for any } n. \text{ From Eq. (6), } \sum_{i=1}^{n} P(B_i) \text{ is convergent because it is increasing and the limit is } \sum_{n=1}^{\infty} P(B_n).
\]
\[ \lim_{n \to \infty} P(A_n) = \sum_{i=1}^{\infty} P(B_n), \text{ but } \sum_{i=1}^{\infty} P(B_n) = P(\bigcup_{n=1}^{\infty} A_n). \]

(4) If \( A_1, \ldots, A_n \) are disjoint, then \( P(\bigcup_{i=1}^{n} A_n) = \sum_{i=1}^{n} P(A_n). \)

Proof: Assume \( A_{n+1}, A_{n+2}, \ldots = \emptyset. \)

(5) If \( A_1 \supset A_2 \supset A_3 \supset \ldots \), then \( \lim_{n \to \infty} P(A_n) = P(\bigcap_{n=1}^{\infty} A_n). \)

Exercise 3. Prove Property (5).

Example 13. Dart Problem: \((\Omega, \mathcal{F})\). Let us define \( P(\{ \text{"miss"}\}) = 0.5. \) If \( A \in \mathcal{F} \) and \( A \subset D \), then \( P(A) = \frac{\text{area}(A)}{2\pi}. \) We can check that \( P \) is a probability measure, e.g., \( P(\{ \text{"miss"}\} \cup D) = P(\{ \text{"miss"}\}) + P(D) = 1/2 + 1/2 = 1. \) We can check the other property as well. Take \( E_n = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1 - 1/n\}. \) \( E_1 \subset E_2 \subset E_3 \ldots. \) We can use Property (3): \( \lim_{n \to \infty} P(E_n) = P(\{ (x,y) \in \mathbb{R}^2 : x^2 + y^2 < 1 \}) = 1/2. \) Note that \((x,y) \in \mathcal{F} \) and \((x,y) \subset D. \)

Reverse: Recall the equivalent statements: We want to prove that if \( X^{-1}((\infty, r)) \in \mathcal{F} \) for any \( r \), then \( X^{-1}((\infty, r]) \in \mathcal{F}. \) Trick: \( (-\infty, r] = \bigcap_{n=1}^{\infty} (-\infty, r+1/n). \) \( X^{-1}((\infty, r]) = X^{-1}(\bigcap_{n=1}^{\infty} (-\infty, r+1/n)) = \bigcap_{n=1}^{\infty} X^{-1}((\infty, r+1/n)) \Rightarrow \) the intersection is also an event \( \in \mathcal{F}. \)

5.4. Expectation. \((\Omega, \mathcal{F}, P) : \) probability space. \( X \) is a non-negative r.v. (\( \sigma(x) \)). We define \( E[X] \) (or the average of \( X \)) as follows:

\[ X^{-1}((a_k, a_{k+1})) = t \]

Figure 11.

6.1. **Expectation.** \((\Omega, \mathcal{F}, P)\): probability space (measure space). Random variable: \(X : \Omega \to [0, 1]\) and \(X\) is \(\mathcal{F}\)-measurable. In particular, \(X = 1\) \(\mathcal{F}\)-measurable. Suppose that \(x > 0\).

![Figure 12](image1)

**Figure 12.**

Let’s form (for each \(n = 1, 2, \ldots\)) \(S_n = \sum_{i=1}^{\infty} \frac{i}{2^n} P(\{\omega \in \Omega : \frac{i}{2^n} < X(\omega) \leq \frac{i+1}{2^n}\})\).

\(\{\omega \in \Omega : a < X(\omega) \leq b\}\) is an event = \(X^{-1}((a, b]) = X^{-1}((\infty, b]) \setminus X^{-1}((\infty, a]) \Rightarrow\) difference is an event. We will show that \(S_1 \leq S_2 \leq S_3 \leq \ldots\). In which case, \(S_n\) is either convergent or \(S_n \not\to \infty\). We define \(E[X] = \lim_{n \to \infty} S_n\): the expectation of \(x\) with respect to probability \(P\). If \(E[X] < \infty\), then we say that \(X\) is integrable. Otherwise, \(X\) is not integrable.

Now, suppose that \(X\) is arbitrary. Then, we can always write: \(X = X^+ - X^-\), where \(X^+ = \max(X, 0)\) and \(X^- = \max(-X, 0)\). Note that both \(X^+\) and \(X^-\) are non-negative r.v.’s. \(\therefore\) The previous expectation stuff applies. In particular, we say \(X\) is integrable if \(E[X^+] < \infty\) and \(E[X^-] < \infty\). In this case, we write \(I\)

![Figure 13](image2)

**Figure 13.**

Show that \(S_n \not\to\). \(S_n = \sum_{i=1}^{\infty} \frac{i}{2^n} P(\frac{i}{2^n} < x \leq \frac{i+1}{2^n}) = \sum_{i=1}^{2i+1} \frac{2i}{2^n} P(\frac{2i}{2^n} < x \leq \frac{2i+2}{2^n})\).

Now, \(\oplus = P(\frac{2i}{2^n} < x \leq \frac{2i+2}{2^n}) + P(\frac{2i+1}{2^n} < x \leq \frac{2i+2}{2^n})\).

\(\therefore S_n = \sum_{i=1}^{\infty} \frac{2i}{2^n}(a_{2i,n+1} + a_{2i+1,n+1})\)
\[ \sum_{i=1}^{\infty} \frac{2^i}{2^{i+1}} a_{2i,n+1} + \sum_{i=1}^{\infty} \frac{2^i}{2^{i+1}} a_{2i+1,n+1} \leq \sum_{i=1}^{\infty} \frac{2^{i+1}}{2^{i+1}} a_{2i,n+1} + \sum_{i=1}^{\infty} \frac{2^{i+1}}{2^{i+1}} a_{2i+1,n+1} \leq \sum_{j=1}^{\infty} a_{j,n+1} = S_{n+1} \]

\[ \therefore \lim_{n \to \infty} S_n \text{ makes sense.} \]

**Remark:** \( E[|X|] = E[X^+] + E[X^-] \).

### 6.2. Distributions.

\((\Omega, \mathcal{F}, P)\); \( x \) is a r.v. We define the set function \( \mu_x : B \to [0,1] \) by the rule:

\[ \mu_x(B) = P\{x \in B\} = P(X^{-1}(B)), B \in \mathcal{B}. \]

Since \( x \) is a r.v., \( X^{-1}(B) \in \mathcal{F} \). \( \mu_x \) is called the *distribution* induced by \( x \).

**Exercise 4.** Show that \( \mu_x \) is a probability measure on \( \mathcal{B} \) (Borel set). Check that \( \mathcal{R} \) is in it: \( \mu_x(\mathcal{R}) = P(X^{-1}(\mathcal{R})) = P(\Omega) = 1 \). Check \( \mu_x(\bigcup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} \mu_x(B_n) \) if \( B_i \cap B_j = \emptyset, i \neq j \).

**Hint:** Establish the connection between \( X^{-1}(\bigcup_{n=1}^{\infty} B_n) \) and \( \bigcup_{n=1}^{\infty} X^{-1}(B_n) \).

### 6.3. Distribution Function.

\( X \) is a r.v. We need to find the distribution function \( F_x(x) \) as (function of real variable \( x \)) \( F_x(x) = \mu_x((-\infty, x]) = P\{-\infty < x \leq x\} \).

**Properties of \( F_x \):**

1. Non-decreasing.
2. \( \lim_{x \to -\infty} F_x(x) = 1, \lim_{x \to -\infty} F_x(x) = 0.\)

**Proof:** \( \bigcup_{n=1}^{\infty} \{ -\infty < x \leq n \} = \Omega \). \( P(\bigcup_{n=1}^{\infty} \{ -\infty < x \leq n \}) = P(\Omega) = 1 \). Left hand side is: \( \lim_{n \to \infty} P\{ -\infty < x \leq n \} = \lim_{n \to \infty} F_x(n) \). We proved: \( \lim_{n \to \infty} F_x(n) = 1 \).

**Exercise 5.** Show that \( \lim_{n \to \infty} F_x(-n) = 0 \). **Hint:** Use the result of HW#2.
6.4. **Density Functions.** If $F_x(\cdot)$ is differentiable, i.e., $\frac{dF_x(x)}{dx} = f_x(x)$, then $E[X] = \int_{-\infty}^{\infty} xf_x(x)dx$.

Why? $E[X] = \lim_{n \to \infty} S_n = \lim_{n \to \infty} \sum_{i=1}^{\infty} \frac{1}{2^n} (F_x (\frac{i+1}{2^n}) - F_x (\frac{i}{2^n})) \approx \lim_{n \to \infty} \sum_{i=1}^{\infty} \frac{1}{2^n} f_x (\frac{i}{2^n}) 2^{-n}$.

But $\frac{1}{2^n} \sum_{i=1}^{\infty} \frac{1}{2^n} f_x (\frac{i}{2^n})$ converges to the Riemann Integral: $\int_{0}^{\infty} xf_x(x)dx$. Notation: We write $E[X] = \int X(\omega)dP(\omega) = \int xP(dx)$: Lebesgue integral of $x$ w.r.t. $P$.

Also, $S_n = \sum_{i=1}^{\infty} \frac{1}{2^n} P(\frac{i}{2^n} < x \leq \frac{i+1}{2^n}) = \sum_{i=1}^{\infty} \frac{1}{2^n} \mu_x ((\frac{i}{2^n}, \frac{i+1}{2^n}))$. We can write $E[X] = \int x\mu_x(dx) = \int xP(dx)$: Lebesgue integral of $x$ w.r.t. $P$.

$(\Omega, \mathcal{F}, P)$. $x$ is a r.v. $E[g(x)] = \lim_{n \to \infty} \sum_{i=1}^{\infty} g(\frac{i}{2^n}) a_{i,n}$. If $E[X^2] < \infty$, we say that $x$ is square integrable.
7. Lecture 5: 09/10/2003

7.1. Elementary Hilbert Space Theory. A complex vector space \( H \) is called an inner product space if \( \forall x, y \in H \), we have a scaler \( < x, y > \) so-called "inner product" such as the following rules are satisfied:

\[
\begin{align*}
(1) & \quad < x, y > = \langle \bar{x}, y \rangle, \forall x, y \in H \\
(2) & \quad < x + y, z > = < x, z > + < y, z >, \forall x, y, z \in H \\
(3) & \quad < \alpha x, y > = \alpha < x, y >, \forall x, y \in H, \alpha \in \mathbb{C} \\
& \quad < x, \alpha y > = \bar{\alpha} < x, y > \\
(4) & \quad < x, x > \geq 0, \forall x \in H \\
& \quad < x, x > = 0 \iff x = 0
\end{align*}
\]

7.2. Schwarz Inequality. Properties Rules (1)-(4) imply that:

\[
| < x, y > | \leq ||x|| ||y||, \forall x, y \in H
\]

Proof: Assume \( A = ||x||^2 \) and \( C = ||y||^2 \) and \( B = || < x, y > \|. \) \( x - \alpha y, x - \alpha y > \geq 0, r \in \mathbb{R}, \alpha \in \mathbb{D}. \) \( < x, x - \alpha y > - < \alpha y, x - \alpha y > \geq 0. \) \( \odot = ||x^2|| - r\alpha < x, y > - r\bar{\alpha} < x, y > + r^2||y^2|| \geq 0, \forall r \in \mathbb{R} \) and \( \forall \alpha \in \mathbb{D}. \) Choose: \( \alpha : \alpha < y, x > = | < x, y > |. \) Note that \( \alpha = \frac{<x,y>}{||<x,y>||} \) and \( |\alpha| = \frac{<x,y>}{||<x,y>||} = 1. \) \( \odot \Rightarrow ||x^2|| - 2r| < x, y > | + r^2||y^2|| \geq 0 \Rightarrow A - 2rB + r^2C \geq 0, r \in \mathbb{R}. \)

\[
r_{1,2} = \frac{2B \pm \sqrt{B^2 - 4AC}}{2C} = \frac{B \pm \sqrt{B^2 - AC}}{C}, \quad B^2 - AC \geq 0. \quad B^2 \geq AC \Rightarrow | < x, y > |^2 \leq ||x||^2 ||y||^2 \Rightarrow | < x, y > | \geq ||x|| ||y||.
\]

7.3. Triangle Inequality. It follows from Schwarz Inequality: \( ||x + y|| \leq ||x|| + ||y||, \forall x, y \in H. \) It follows from the Triangle Inequality that: \( ||x - z|| \leq ||x - y|| + ||y - z||. \) If we look at the distance between the two vectors \( x, y \in H \) as: \( d(x, y) = ||x - y||, \) then we define \( H \) with the metric \( d \) as a metric space \((H, d)\). \( d : H \times H \rightarrow \mathbb{R^+}. \) Now, if \( H \) is complete, then \( H \) is called a Hilbert Space. Complete space \( H \) means any Cauchy sequence in \( H \) converges to a point that lies in \( H. \) Cauchy sequence: \( \{y_n\}_{n=1}^{\infty} \) is called a Cauchy sequence if for any \( \epsilon > 0, \exists \) a large number \( N \) such that \( ||y_n - y_m|| \leq \epsilon \) for \( n, m > N. \)

Exercise 6. Prove the Triangle Inequality. Hint: \( < x + y, x + y > = . . . . \)

\[\text{Figure 15.}\]
7.4. Convex Sets. A set $E$ in a vector space $\mathcal{V}$ is said to be a convex set if for any $x, y \in E$, and $t \in (0, 1)$, the following point $Z_t = tx + (1 - t)y \in E$. In other words, the line segment between $x$ and $y$ lies in $E$. $E + x = \{y + x : y \in E\}$ is a convex set.

7.5. Orthogonality. If $<x, y> = 0$, then we say that $x$ and $y$ are orthogonal. We write $x \perp y$. $\perp$ is a symmetric relation: $x \perp y \iff y \perp x$.

Pick a vector $x \in H$, then find all vectors in $H$ that are orthogonal to $x$. We write it as:

$x^\perp = \{y \in H : <x, y> = 0\}$ and $x^\perp$ is a closed subspace.

Let $M$ be a subspace in $H (M \subset H)$. Define $M^\perp$. $M^\perp = \bigcap_{x \in M} x^\perp, \forall x \in M$.

**Theorem 3.** Any non-empty, closed convex set $E \subset H$ contains a unique element with smallest norm. $\exists x_o \in E : ||x_o|| < \infty$.

Proof: “Parallelogram I.

$x \in E$.

\[ \begin{figure}[h] \centering \begin{tikzpicture} [scale=0.5] \draw[->, thick] (0,0) -- (6,0); \draw[->, thick] (0,0) -- (0,6); \draw[->, thick] (0,0) -- (3,3); \draw[->, thick] (3,3) -- (6,0); \draw[->, thick] (3,3) -- (0,6); \end{tikzpicture} \caption{Figure 16.} \end{figure} \]

Uniqueness: Replace $x$ by $x/2$ and $y$ by $y/2$. In the parallelogram law: $||x/2 + y/2||^2 - ||x/2 - y/2||^2 = \frac{1}{2}||x||^2 + \frac{1}{2}||y||^2$. $||x - y||^2 = 2||x||^2 + 2||y||^2 - 4|\frac{x+y}{2}|^2$. Assume that $\frac{x+y}{2} \in E$ ($E$ is convex). $||\frac{x+y}{2}||^2 = \delta^2$.

\[ ||x - y||^2 = 2||x||^2 + 2||y||^2 - 4\delta^2. \]

Assume $||x|| = ||y|| = \delta$. Substitute in Eq. (6) $\Rightarrow ||x - y||^2 = 2\delta^2 + 2\delta^2 - 4\delta^2 = 0 \Rightarrow x = y \Rightarrow$ there is a unique element.

Existence: Assume $\{y_n\}_{n=1}^\infty$ such that $\lim_{n \to \infty} ||y_n|| = \delta$. $||x - y||^2 = 2||x||^2 + 2||y||^2 - 4\delta^2$.
Replace $x$ by $y_n$ and $y$ by $y_m$. Then $||y_n - y_m||^2 = 2||y_n||^2 + 2||y_m||^2 - 4\delta^2$. $\lim_{n, m \to \infty} ||y_n - y_m||^2 = 2\lim_{n \to \infty} ||y_n||^2 + 2\lim_{m \to \infty} ||y_m||^2 - 4\delta^2 \Rightarrow \lim_{n, m \to \infty} ||y_n - y_m||^2 = 2\delta^2 + 2\delta^2 - 4\delta^2 = 0$. So, $\{y_n\}_{n=1}^\infty$ is a Cauchy sequence. Since $H$ is complete then $\exists x_o \in H : y_n \to x_o$.

**Example 14.** For any fixed $n$, the set $\mathcal{D}^n$ of all $n$-tuples $x = (x_1, x_2, \ldots, x_n), x_i \in D$ is a Hilbert Space, where $<x, y> = \sum_{i=1}^{n} x_i \bar{y}_i, Y = (y_1, y_2, \ldots, y_n)$.

**Example 15.** $L_2[a, b] = \{ f(x) : f^{(b)}(a) |f(x)|^2dx < \infty, x \in [a, b] \}$ is a Hilbert Space, where $<f, g> = \int_a^b f(x)\overline{g(x)}dx$. $||f|| = \sqrt{<f, f>} = [\int_a^b |f|^2dx]^{1/2} = ||f||_2$. 


Exercise 7. Show that the Schwarz Inequality is an equality if $y = \alpha x$, i.e., $| < x, \alpha x > | = ||x|| \cdot ||\alpha x||$.

8.1. Triangle Inequality. $< x + y, x + y >= < x, x > + < y, y > + < x, y > + < y, x > = ||x||^2 + ||y||^2 + < x, y > + < x, y >^*$. But, if $c \in D$, then $c + c^* \leq 2|c|$. $||x + y||^2 = ||x||^2 + ||y||^2 + 2 < x, y > = (||x|| + ||y||)^2$. \( \therefore ||x + y|| \leq ||x|| + ||y|| \). Equality holds if $< x, y >= 0$ ($x \perp y$).

8.2. Notion of a Closed Set. We have a norm, i.e., we say $||x||$ is a norm on $H$ if:

1. $||x|| \geq 0$.
2. $||x|| = 0$ only if $x = 0$.
3. $||x + y|| \leq ||x|| + ||y||$.
4. $||\alpha x|| = |\alpha| ||x||$, $\alpha \in D$.

Exercise 8. Show that $||x|| = < x, x >^{1/2}$ is a norm on $H$.

We say $x_n \to x$, whenever $x_n \in H$ and $x \in H$ if $\lim_{n \to \infty} ||x_n - x|| = 0$ or $||x_n - x|| \to 0$ as $n \to \infty$. Suppose that $M \subset H$. We say that $M$ is closed if it happens that $x_n \to x \in H$, then necessarily $x \in M$. $H$ is called complete whenever $x_n, x_m$ have the property that $||x_n - x_m|| \to 0$ as $m, n$ tend to $\infty$. Then, $x_n$ is a convergent sequence, i.e., $\exists x_o \in H$ for which $x_n \to x_o$. $H$ is complete if every Cauchy sequence is convergent.

Fact: If $H$ is complete, then it is closed.

Recall: If $M$ is a clo

$M$ contains a unique

\[
\mathcal{H} = \mathbb{R}^2
\]

\[
\text{element of minimal norm}
\]

\[
\mathcal{M}
\]

\[\text{Figure 17. A line is closed. } \mathcal{M} \text{ is convex and closed.}\]
Theorem 4. Projection onto closed subsets. If $M \subset H$, $H$ is in Hilbert Space, $M$ is closed. Then for every $x \in H$, $\exists$ a decomposition $x = P_x + Q_x$, where $P_x \in M$, $Q_x \in M^\perp$ ($M^\perp = \{ y \in H | (x, y) = 0 \}$) and

1. The decomposition is unique.
2. $||x||^2 = ||P_x||^2 + ||Q_x||^2$.
3. If $M \neq H$, then $\exists y \in H$, $y \notin M$ such that $y \perp M$.
4. $P_x$ is the nearest point in $M$ to $x$. $Q_x$ is the nearest point in $M^\perp$ to $x$. $P_x$ is called the projection of $x$ into $M$.

\[ H = \mathbb{R}^2 \]
\[ M = \mathbb{R} \]
\[ Q_x \]
\[ P_x \]
\[ M \]

Figure 18.

8.3. Application. ($\Omega, \mathcal{F}, P$). Let $\mathcal{D} \subset \mathcal{F}$ (e.g., $\mathcal{D}$ could be $\sigma(x)$ if $x$ is a r.v. defined on ($\Omega, \mathcal{F}$). Let $L_2(P)$ denote all the r.v.’s ($\mathcal{F}$-measurable) that are square integrable, i.e., if $x \in L_2(P)$, then $E[|x|^2] < \infty$.

Claim: $L_2(P)$ is a vector space. Check closure under addition. Suppose $X, Y \in L_2(P)$, we need to check that $E[|X + Y|^2] < \infty$ in which case $X + Y \in L_2(P)$. What is an inner product here? We define $<X, Y> = E[XY]$ (think of $XY$ as a new variable $Z$). Also, $<X, X> = |X|^2 = E[X^2]$. We can check that $<X, Y>$ is an inner product. We need to show that: $||X + Y|| < \infty$ (finite) $\Rightarrow E[|X + Y|^2]^{1/2}$. $E[|X + Y|^2] = ||X + Y||^2 \leq ||X||^2 + ||Y||^2 \Rightarrow < \infty$. Also, let’s do closure under scalar multiplication. If $x \in L_2(P)$, then $||\alpha x|| = E[|\alpha x|^2]^{1/2} = |\alpha|E[|x|^2]^{1/2} = |\alpha| ||x|| < \infty$.

Let $L_2(\mathcal{D})$ be the collection of all $\mathcal{D}$-measurable square integrable r.v.’s, i.e. if $x \in L_2(\mathcal{D})$ then,

1. $x$ is $\mathcal{D}$-measurable.
2. $E[|x|^2] < \infty$.

Then, $L_2(\mathcal{D})$ is also a vector space. Note that $L_2(\mathcal{D}) \subset L_2(P)$ and since $L_2(\mathcal{D})$ is itself a vector space $L_2(\mathcal{D})$ is a subspace of $L_2(P)$. We have the following:

1. $L_2(P)$ is a vector space (its also complete) so its a Hilbert Space.
2. $L_2(\mathcal{D})$ is a closed subspace of $L_2(P)$. 
Think of $L_2(P)$ as $\mathcal{H}$. Think of $L_2(\mathcal{D})$ as $\mathcal{M}$. According to the projection theorem, if $x \in L_2(P)$, then we can write $x$ as $X = PX + QX$, where $PX \in \mathcal{M}$ and $QX \in \mathcal{M}^\perp$. We define $E[X|\mathcal{D}] \triangleq PX$.

First, note that $E[X|\mathcal{D}]$ is a $\mathcal{D}$-measurable r.v.

**Exercise 9.** (1) Show that if $x$ is a $\mathcal{D}$-measurable r.v. then so is $aX$, $a \in \mathbb{R}$. (2) Show that if $X$ and $Y$ are $\mathcal{D}$-measurable, so is $X + Y$.

**Special Case:** $\mathcal{D} = \sigma(Y), Y \in L_2(P)$. Then, we can still talk about $E[x|\sigma(Y)]$: is a square integrable $\sigma(Y)$-measurable r.v.

**Theorem 5.** If $Z$ is a $\sigma(Y)$-measurable r.v. then we can write $Z = h(Y)$, where $h$ is a Borel-measurable function, meaning: $h^{-1}(B) \in \mathfrak{B}$ (or all $B \in \mathfrak{B}$). $E[X|\sigma(Y)]$ is a function of $Y$ and we write it as $E[X|Y]$. 
Theorem 6. Projection Theorem: \( \mathcal{H} \): Hilbert Space. \( \mathcal{M} \): Closed subset in \( \mathcal{H} \).

(1) Every \( x \in \mathcal{H} \) has a unique decomposition. \( x = Px + Qx \), where \( Px \in \mathcal{M} \) and \( Qx \in \mathcal{M}^\perp \).

(2) \( Px \) is the nearest point in \( \mathcal{M} \) to \( x \). \( Qx \) is the nearest point in \( \mathcal{M}^\perp \) to \( x \).

(3) If we think of \( Px \) as mapping \( x \) to \( Px \), then \( P \) is linear. The same can be said about \( Qx \).

(4) \( ||x||^2 = ||Px||^2 + ||Qx||^2 \).

Proof: Uniqueness. Suppose that we have \( x = x' + y' \) and \( x = x'' + y'' \). Then, \( x' + y' = x'' + y'' \), \( x' - x'' = y'' - y' \). \( \therefore x' - x'' = y'' - y' = 0 \) since the same element \( \in \mathcal{M} \) and \( \in \mathcal{M}^\perp \).

\( \therefore x' = x'', \ y' = y'' \) unique.

Consider \( \mathcal{M} + x = \{ x + y : y \in \mathcal{M} \} \). Claim: \( \mathcal{M} + x \) is closed. \( \mathcal{M} \) is closed, so, if \( x_n \in \mathcal{M} \) and as \( n \to \infty \), \( x_n \to x_o \) (\( ||x_n - x_o|| \to 0 \)) then \( x_o \to \mathcal{M} \). Pick a convergent sequence in \( x + \mathcal{M} \), call it \( Z_n \). \( Z_n = x + y_n \), \( y_n \in \mathcal{M} \). Since \( Z_n \) is convergent, so is \( y_n \), but the limit of \( y_n \) is in \( \mathcal{M} \) since \( \mathcal{M} \) is closed.
We show that $x + M$ is convex. Pick $x_1$ and $x_2 \in x + M$. We need to show that for any $0 \leq \alpha \leq 1$, $\alpha x_1 + (1 - \alpha)x_2 \in x + M$. But $x_1 = x + y_1$, $y_1 \in M$ and $x_2 = x + y_2$, $y_2 \in M$. $\alpha x_1 = \alpha x + \alpha y_1 \rightarrow \alpha x_1 + (1 - \alpha)x_2 = x + \alpha y_1 + (1 - \alpha)y_2 \rightarrow x + M \in M$ (since $M$ is convex). $(1 - \alpha)x_2 = (1 - \alpha)x + (1 - \alpha)y_2 \in x + M$. Let $x = x_2 = (1 - \alpha)x + (1 - \alpha)y_2 \in x + M$. By the earlier theorem, $3$ a member in $x + M$ of smallest norm. Call it $Qx$. Let $P_x = x - Qx$. Note that $P_x \in M$. We need to show that $Qx \in M$. Namely, $< Qx, y > = 0$, $\forall y \in M$. Call $Qx = z$. $||z|| \leq ||y||$, $\forall y \in M + x$. Pick $y = z - 2y$, where $y \in M$, $||y|| = 1$. $||z||^2 \leq ||z - 2y||^2 = < z - 2y, z - 2y >. 0 \leq -\alpha < y, z > -\alpha < z, y > + ||\alpha||^2$. Pick $\alpha = < z, y >$. We obtain $0 \leq | < z, y >|^2$. This can hold only if $< z, y > = 0$, i.e., $z$ is orthogonal to every $y \in M$. This is a member in $M$ of smallest norm. Thus $Qx \in M$.  

9.1. Minimum Distance Properties. Show that $P_x$ is the nearest point in $M$ to $x$. $y \in M$, $||x - y||^2 = ||P_x + Qx - y||^2 = ||Qx + (P_x - y)||^2 = ||Qx||^2 + ||P_x - y||^2$: Pythagoras. $P_x$ is the member of $M$ which is nearest to $x$. The $Qx$ case is shown similarly.

9.2. Linearity. Take $x, y \in \mathcal{H}$. $x = P_x + Qx$. $y = P_y + Qy$. $(ax + by) = P(ax + by) + Q(ax + by) = aP_x + aQx + bPy + QQy$. $P(ax + by) - aP_x - bP_y = -Q(ax + by) + aQx + bQy \rightarrow$ The only vector that satisfies the equation is $0$. $P(ax + by) = aP_x + bP_y$. $Q(ax + by) = aQx + bQy$.

Comment: If $x \in M$, then $Qx = 0$ and $x = P_x$. $(\Omega, \mathcal{F}, P)$. $x$ is a r.v. $\mathcal{D} \subset \mathcal{F}$. $L_2(\mathcal{P}) = \{Y : E[|Y|^2] < \infty\}$. $L_2(\mathcal{P})$ is a vector space. $L_2(\mathcal{P})$ is closed. Why? $x_n \rightarrow x_o$ (this means that $E[|x_n - x_o|^2] \rightarrow 0$ as $n \rightarrow \infty$). $x_n \in L_2(\mathcal{P})$. $E[x_n] < \infty$. $E[x_n^2] = E[(x_n - x_o + x_o)^2] = E[(x_n - x_o)^2] + E[x_o^2] + 2E[x_n(x_n - x_o)] \leq 2E[x_n^2]^{1/2}E[(x_n - x_o)^2]^{1/2} < \infty$. $L_2(\mathcal{P})$ is closed inner product space: $< x, y >= E[XY]$. $L_2(\mathcal{D}) = \{\mathcal{D}$-measurable $Y : E[|Y|^2] < \infty\}$. $L_2(\mathcal{D})$ is also closed (same reason). Let $M = L_2(\mathcal{D})$ and $H = L_2(\mathcal{P})$. We apply the projection theorem to $x \in L_2(\mathcal{P})$. $x = P_x + Qx$, where $P_x \in L_2(\mathcal{D})$, $Qx \in L_2(\mathcal{D})$. We have the property that: $||x - P_x|| \leq ||x - y|| \forall y \in L_2(\mathcal{D})$. We call $P_x$ the conditional expectation of $x$ given $\mathcal{D}$. $P_x = E[x|\mathcal{D}]$.


(1) Call $E[X|\mathcal{D}] = Z$. $E[XY] = E[ZY], Y \in L_2(\mathcal{D})$. Interpretation: $Z$ contains all the information that $X$ contains relevant to $Y$.

Proof: $E[XY] = E[(PX + QY)Y] = E[(PX)Y] + E[(QX)Y] = E[ZY]$. Conversely, if a r.v. in $L_2(\mathcal{D})$ has the property that $E[ZY] = E[XY]\forall Y \in L_2(\mathcal{D})$ then $Z = E[X|\mathcal{D}]$.

Alternative definition of $E[X|\mathcal{D}]$. $z = E[X|\mathcal{D}]$ if $E[ZY] = E[XY]\forall Y \in L_2(\mathcal{D})$.

(2) $E[a|\mathcal{D}] = a$. We need to show that $E[aY] = E[aY]$.

(3) $E[X] = E[E[X|\mathcal{D}]]$. $Z = E[X|\mathcal{D}]$. $E[ZY] = E[XY], \forall Y \in L_2(\mathcal{D})$, for $Y = 1$ (note that 1 is $\{\Omega, \emptyset\}$ measurable, so it is also $\mathcal{D}$-measurable) $\rightarrow E[X] = E[Z]$.

(4) $E[Y|\mathcal{D}] = Y$ if $Y \in L_2(\mathcal{D})$ (e.g., $E[X|X] = X$). Check: $E[Y\omega] = E[Y\omega], \omega \in L_2(\mathcal{D})$. 

ECE 541 33
(5) $E[XY|\mathcal{D}] = YE[X|\mathcal{D}]$ if $Y \in L_2(\mathcal{D})(E[XY|Y] = YE[X|Y]$.

(6) Let $\mathcal{D}_1 \subset \mathcal{D}_2 \subset \mathcal{F}$. $E[X|\mathcal{D}_1] = E[E[X|\mathcal{D}_2]|\mathcal{D}_1]$. Also, $E[X|\mathcal{D}_1] = E[E[X|\mathcal{D}_1]|\mathcal{D}_2]$. Take $\mathcal{D} = \{\Omega, \emptyset\}$. Then $E[X|\mathcal{D}] = E[X]$. $E[X|\mathcal{D}]$ is $\mathcal{D}$-measurable. $E[X|\mathcal{D}]$ must be a cte if $\mathcal{D} = \{\Omega, \emptyset\}$.

Let $\mathcal{D} = \sigma(Y), Y \in L_2(P)$. Then $E[X|\mathcal{D}] = Z = h(Y)$.

**Theorem 7.** If $Z$ is $\mathcal{D}$-measurable, then there exists a function $h$ such that $Z = h(Y)$.

In this case, we write $Z = E[X|Y]$. 

10. Lecture 8: 09/22/03

If $Z = E[X|\mathcal{D}], \mathcal{D} \subset \mathcal{F}$, then

(7) $E[ZY] = E[XY], Y \in L_2(\mathcal{D})$.

We use Eq. (7) as the defining property for $E[X|Y]$. To be able to do so, we need to show: If $Z$, a $\mathcal{D}$-measurable r.v., has the property that $E[XY] = E[ZY]$, then $Z$ must be $= E[X|\mathcal{D}] (= PX)$.

We need to show that $Z = PX$. Note that $E[XY - ZY] = 0, \forall Y \in L_2(\mathcal{D})$. \therefore $E[Y(X - Z)] = 0$. $E[Y(PX + QY - Z)] = 0$. $E[YPX] + E[YQX] - E[YZ] = 0$. $E[Y(Z - PX)] = 0, \forall Y \in L_2(\mathcal{D})$. In particular, take

(8) $Y = \begin{cases} 1, & \text{if } Z - PX \geq 0 \\ -1, & \text{if } Z - PX < 0 \end{cases}$

Claim: $Y \in L_2(\mathcal{D})$. $E[Y(Z - PX)] = E[|Z - PX|] = 0$ by the assumption. \therefore $Z - PX = 0 \rightarrow Z = PX$.

**Property 9.** **Smoothing Property:** $E[X|\mathcal{D}_1] = E[E[X|\mathcal{D}_2]|\mathcal{D}_1]. \mathcal{D}_1 \subset \mathcal{D}_2 \subset \mathcal{F}.

Proof: Need to show: $E[ZY] = E[XY], \forall Y \in L_2(\mathcal{D}_1)$. LHS: $E[YE[E[X|\mathcal{D}_2]|\mathcal{D}_1]]$. Note that $E[E[X|\mathcal{D}_2]|\mathcal{D}_1]$ is $\mathcal{D}_1$-measurable $= E[E[YE[X|\mathcal{D}_2]|\mathcal{D}_1]]$. Note that $Y$ is also $\mathcal{D}_2$-measurable ($\in L_2(\mathcal{D}_2)$) $= E[E[E[X|\mathcal{D}_2]|\mathcal{D}_1]]$. But, $E[E[YX|\mathcal{D}_2]] = E[XY] = E[E[XY]] = E[XY]$.

**Property 10.** If $X$ and $Y$ are independent r.v.’s, then $E[g(x,y)] = E[h(y)],$ where $h(t) = E[g(x,t)]$.


$E[h(Y)] = E[YE[X]|Y]].$ Omit general proof beyond special case.

**Example 16.** **Photon counting:** Energy of photon = $hv$, where $h$ is Planck’s constant and $v$ is the wave frequency in Hertz.

---

![Figure 21.](image-url)
Assume: the height of each pulse is random and the counter integrates the pulse and every other pulse. \( M = \sum_{i=1}^{N} G_i \), where \( G_i \) is the area under the \( i \)th pulse. \( N \) is the number of photons during the detector period \( \tau \). \( N \) and \( G_i \) are random. Assume that \( G_i \) and \( N \) are independent. Goal: Find \( \mathbb{E}[M] \). What do we know? \( P\{M = K\} \) is known (\( M = 0, 1, 2, 3, \ldots \)). \( P\{G_i = K\} \) is known (\( G_i = 0, 1, 2, 3, \ldots \)). \( \mathbb{E}[M] = \mathbb{E}[\mathbb{E}[M|N]] = \mathbb{E}[h(N)] \), where \( h(t) = \mathbb{E}\left[ \sum_{i=1}^{t} G_i \right] = \sum_{i=1}^{t} \mathbb{E}[G_i] \). Let’s assume that \( \mathbb{E}[G_i] = \alpha \) \( \forall i \). \( h(t) = t\alpha \). \( h(N) = \alpha N \). \( \mathbb{E}[M] = \mathbb{E}[h(N)] = \alpha \mathbb{E}[N] \). What is independent is the collection \( \{G_i\} \) and \( N \). What we really needed: Let \( \bar{w} = (Y_1, Y_2, \ldots) \). Let \( \bar{w} \) and \( x \) be independent. Then \( \mathbb{E}[g(x, \bar{w})] = \mathbb{E}[h(x)] \), where \( h(t) = \mathbb{E}[g(t, \bar{w})] \). In our example, \( w \) was \( \bar{w} = [G_1, G_2, \ldots] \). \( x = N \). What is the variance of \( M \)? We know \( \sigma_n^2 \) and \( \sigma_{G_i}^2 \). \( G_i \)’s are independent of each other. Let’s calculate \( \mathbb{E}[M^2] \). \( \mathbb{E}[M^2] = \mathbb{E}[\mathbb{E}[M^2|N]] = \mathbb{E}[h_2(N)] \), where \( h_2(t) = \mathbb{E}[\left( \sum_{i=1}^{t} G_i \right)^2] = \mathbb{E}[\sum_{i=1}^{t} \sum_{j=1}^{t} G_i G_j] = \mathbb{E}[\sum_{i \neq j} \sum G_i G_j] + \mathbb{E}[\sum_{i \neq j} \sum G_i^2] = (t^2 - t)\alpha^2 + t(\alpha^2 + \sigma_{G_i}^2) \equiv h_2(t) \).

\[ E[h_2(N)] = \mathbb{E}[\left( N^2 - N \right)] + (\alpha^2 + \sigma_{G_i}^2) \mathbb{E}[N] = \alpha^2(\sigma_N^2 + N^2 - \bar{N}) + (\alpha^2 + \sigma_{G_i}^2) \bar{N}, \] where \( \bar{N} = \mathbb{E}[N] \). \( \sigma_M = \alpha^2(\sigma_N^2 + N^2 - \bar{N}) + (\alpha^2 + \sigma_{G_i}^2) \bar{N} \). \( \sigma_M - \alpha^2 \bar{N}^2 = \alpha^2(\sigma_N^2 - \bar{N}) + (\alpha^2 + \sigma_{G_i}^2) \bar{N} \).
11. Lecture 9: 09/29/03

11.1. Applications of Conditional Expectations: Problem: We flip a H-T coin successively. Suppose $P\{H\} = p$ and $P\{T\} = q = 1 - p$. Coin flips are independent of each other. $Y_1 = \min\{i \geq 1 : x_1, \ldots, x_{i-1} = 0, x_i = 1\}$, where

\begin{equation}
    x_i = \begin{cases} 
        1, & \text{$i$th flip is a “H”} \\
        0, & \text{$i$th flip is a “T”}
    \end{cases}
\end{equation}

$x_i$’s are independent of each other. $Y_1$ is the flip index at which we see a head for the first time. $E[Y_1]$. . .

NOTE: THIS IS AN INTENTIONALLY SHORT LECTURE.
Experiment: Toss a coin repeatedly, infinitely many times and observe a sequence of heads and tails. Random variables: We define \( T_n, n = 1, 2, 3, \ldots; \omega = (\omega_1, \omega_2, \ldots) \).

\[
T_n = \begin{cases} 
1, & \omega_n = H \\
0, & \omega_n = T 
\end{cases}
\]

where \( Y_1 = \min\{i : x_i = 1\} \). More generally, \( Y_k = \min\{i : T_i = T_{i+1} = \ldots = T_{i+k-1} = 1\} \) for each \( Y_k, Y_k \) is a r.v. on \( \Omega = \{(\omega_1, \omega_2, \ldots) : \omega_i \in \{H,T\}\} \). \( Y_k = E[Y_k] \).

Special case: \( k = 1 \), Range \( Y_1 \)? \( R(Y_1) = \{1, 2, 3, 4, \ldots\} \). Assume \( P(H) = p \) and \( q = 1 - p \) (\( P(T) \)).

Then, assuming coins are independently flipped:

\[
P(Y_1 = 1) = p. \\
P(Y_1 = 2) = qp. \\
P(Y_1 = 3) = q^2p
\]

Define \( g(q) = \sum_{i=1}^{\infty} iq^i = (-1 + \frac{1}{1-q}) = -1 + \frac{1}{1-q} \) and \( g'(q) = \sum_{i=1}^{\infty} i^2 q^{i-1} = \frac{1}{q} \sum_{i=1}^{\infty} iq^i \).

Claim: \( Y_k = h(Y_{k-1}, T(Y_{k-1}+k-1), T(Y_{k-1}+k), \ldots) \). Note that \( Y_{k-1} \) and \( \{T_{Y_{k-1}+k-1}, T_{Y_{k-1}+k}, \ldots\} \) are independent. \( E[Y_k] = E[h(Y_{k-1}, T(Y_{k-1}+k-1), T(Y_{k-1}+k)), \ldots] = E[g(Y_{k-1})] \), where \( g(t) = E[h(t, T_{Y_{k-1}+k-1}, \ldots)] \).

Now, we fix \( Y_{k-1} = t \). Consider \( E[h(t, T_{t+k-1}, T_{t+k}, \ldots)] \). Notation: \( A \in \mathcal{F} \). We define \( I_A(\omega) \), a r.v., as follows:

\[
I_A(\omega) = \begin{cases} 
1, & \omega \in A \\
0, & \omega \notin A 
\end{cases}
\]

which we call the indicator function for the event \( A \). Let \( A = \{T_{t+k-1} = 1\} \). \( E[h(t, T_{t+k-1}, \ldots)] = \sum_{i=1}^{\infty} i p^{i-1} (I_A) \).

General Solution: \( y_k = \frac{1}{p} \). If \( p = \frac{1}{2} \), then \( y_2 = 2y_1 + 1 = 5 \).

\[
y_1 = \frac{1}{p}, \quad y_2 = \frac{1}{p} + 1 + \frac{p^2}{q}(1-k) - k - \frac{p}{q}.
\]
Example 17. Example of Property (7) of $E[X|Y]$. In $E[h(X,Y)] = E[g(Y)]$, where $g(t) = E[h(x,t)]$ if $X$ and $Y$ are independent.

Example 18. Counter-example. Let $Y = XZ$ where $Z = 1 + X$. Now, $E[Y] = E[XZ] = E[X(1 + X)] = \bar{x} + \bar{x^2}$. By property (7), $z = t$, $E[XZ] = E[X]t$. $E[E[X|Z]] = E[X]E[Z] = E[X](1 + E[X]) = \bar{x} + \bar{x^2}$. Note that the two are different, i.e., $\bar{x^2} \neq \bar{x^2}$. 
13.1. Independence. \((\Omega, \mathcal{F}, P)\). \(\mathcal{D}_1 \subset \mathcal{D}_2 \subset \mathcal{F}\) We say \(\mathcal{D}_1\) and \(\mathcal{D}_2\) are independent sub-\(\sigma\)-algebras if \(P(A \cap B) = P(A)P(B)\forall A \in \mathcal{D}_1, B \in \mathcal{D}_2\). If \(x\) is \(\mathcal{D}_1\)-measurable, \(Y\) is \(\mathcal{D}_2\)-measurable, we say that \(X\) and \(Y\) are independent. Naturally, if \(X\) and \(Y\) are independent, then \(P\{X \in A, Y \in B\}\) for 
\(A, B \in \mathcal{B} (=P(\{x \in A\} \cap P\{Y \in B\}) = P\{x \in A\}P\{Y \in B\}\).

**Special Case:** Take \(\mathcal{D}_1 = \sigma(x)\) and \(\mathcal{D}_2 = \sigma(y)\). \(X\) and \(Y\) are independent if \(\sigma(x)\) and \(\sigma(y)\) are independent. Claim: \(\mu_{x,y}(A, B) = \mu_x(A)\mu_y(B), \forall A, B \in \mathcal{B}, \) where \(\mu_{x,y}(A, B) = P\{x \in A, y \in B\}\). The proof is clear.

\(\mu_{x,y}\) is called the joint distribution of the r.v.'s \(X\) and \(Y\). If \(X\) and \(Y\) are independent, then \(F_{xy}(x, y) = F_x(x)F_y(y)\). If \(X\) and \(Y\) have densities, then the joint pdf of \(X\) and \(Y\) also factors if \(X\) and \(Y\) are independent.

\[f_{xy}(x, y) = f_x(x)f_y(y).\]

In general, as a function of \(A\), \(\mu_{xy}(A, B)\) is a measure on \(\mathfrak{B}\). As a function of \(B\), \(\mu_{xy}(A, B)\) is a measure on \(\mathfrak{B}\). If \(A = (-\infty, x]\) and \(B = (-\infty, y]\), then \(\mu_{xy}((-\infty, x], (\infty, y]) = P\{X \leq x, Y \leq y\} = F_{xy}(x, y)\).

\[\sum f(x_i)\delta x_i \rightarrow \int_a^b f(t)dt \text{ (Riemann)} = \lim_{\delta y_i \to 0} \sum y_i \mu(y_i < y \leq y_{i+1}) = \int_a^b f(t)d\mu \text{ (Lebesgue measure on } \mathfrak{R}).\]

**Figure 22.**

\[\frac{1}{\sqrt{x}} \not\text{ integrable}\]

**Figure 23.** The inverse of this function looks like \(\frac{1}{\sqrt{x}}\) → it is integrable (finite) but cannot be used in the Riemann Integral.
13.2. **Parameter Estimation.** Suppose that $x$ is a r.v. It has a pdf $f_x(\cdot); f_x(\cdot)$ is parameterized by a parameter $\theta \in \Theta$.

**Example 19.** $x$ is a Gaussian with mean $\theta$ and variance 1. $f_x(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\theta)^2/2\sigma^2}, \Theta$ is arbitrary.

Let $x_1, x_2, x_3, \ldots, x_n$ be independent samples (realizations) of $x$. A function $\hat{\Theta}_n$ defined on $x_1, \ldots, x_n$ is called an estimator of the parameter $\theta$ if the range of $\hat{\Theta}_n$ is in $\Theta$.

Recall $f_x(x) \sim N(\theta, 1)$. To extract $\theta$, we can do:

1. Threshold test.
2. $\frac{1}{n} \sum_{i=1}^{n} x_i$: the fluctuations average to 0.
3. Plug $x_1, \ldots, x_n$ into $f_{x_1}(\cdot)f_{x_2}(\cdot) \ldots f_{x_n}(\cdot)$ (look at the likelihood of the parameters) $\rightarrow$ Maximum Likelihood Estimation (MLE).
4. Find $\hat{\theta}_n(x_1 \ldots x_n)$ for which the mean square error is minimized, i.e., $E[|\hat{\theta}_n(x_1 \ldots x_n) - \theta|^2]$ is minimal $\rightarrow$ Minimum Mean Square Error estimators (MMSE).
5. Linear Estimator
   
   $a_2x_2 + \ldots + a$

   **Figure 25.** We want a small “offset”.
The bias of an estimator $\hat{\theta}_n$ is $b(\hat{\theta}_n) = E[\hat{\theta}_n] - \theta$. The variability of an estimator $\hat{\theta}_n$ is captured by the variance of $\hat{\theta}_n$: $\text{var}(\hat{\theta}_n)$.

For now, let’s consider: $P[\left|\frac{1}{n} \sum_{i=1}^{n} x_i - \theta\right| > \epsilon]$. Assume that $b(\hat{\theta}_n) = 0$ (no offset). $E[\frac{1}{n} \sum_{i=1}^{n} x_i] = \frac{1}{n} \sum_{i=1}^{n} E[x_i] = \frac{1}{n} \sum_{i=1}^{n} \theta = \frac{\theta n}{n} = \theta \rightarrow$ Assumption is correct. Put $Y = \frac{1}{n} \sum_{i=1}^{n} x_i$, $E[Y] = \bar{Y}$. $P[|Y - \bar{Y}| > \epsilon] = E[I_{(|Y - \bar{Y}| > \epsilon)}(\omega)] \leq E[I_{(|Y - \bar{Y}| > \epsilon)}(\omega)\frac{|Y - \bar{Y}|}{\epsilon}] \leq E[\frac{|Y - \bar{Y}|^2}{\epsilon}] = \frac{1}{\epsilon^2}E[|Y - \bar{Y}|^2] = \frac{1}{\epsilon^2}\sigma_Y^2$.

In our case, $Y = \frac{1}{n} \sum_{i=1}^{n} x_i$. $\sigma_Y^2 = \frac{1}{n}$. $\therefore P[\left|\frac{1}{n} \sum_{i=1}^{n} x_i - \theta\right| > \epsilon] \leq \frac{1}{\epsilon^2} \frac{1}{n}$. For any $\epsilon > 0$, if we make $n$ large enough we can drive the above probability to zero.
14. Lecture 12: 10/13/03

14.1. Parameter Estimation

Find a function on \((x_1, x_2, \ldots, x_n)\),
\[ \frac{1}{n} \sum_{i=1}^{n} x_i. \]

\[
E_{\hat{n}} = \frac{1}{n}. \]

By Chebychev, \[ P\{|\bar{x} - E[X]| > \epsilon\} \leq \frac{\text{var}(x)}{\epsilon^2}. \]

General form: If \(\phi\) is an increasing function on \([0, \infty)\), then \[ P\{|\phi(x) - E[\phi(x)]| > \epsilon\} \leq \frac{\text{var}(\phi(x))}{\epsilon^2}. \]

(\(\phi\) is an increasing function on \([0, \infty)\)), \[ E[\hat{\Theta}_n] = \theta. \]

By Chebychev, \[ P\{|\hat{\Theta}_n - \theta| > \epsilon\} \leq \frac{\text{var}(\hat{\Theta}_n)}{\epsilon^2}. \]

Fact (check): \(\text{var}(\hat{\Theta}_n) = \frac{1}{n^2} \sigma^2\). \[ \therefore P\{|\bar{x} - E[\bar{x}]| > \epsilon\} \leq \frac{\sigma^2}{n \epsilon^2}. \]

Claim: \(\lim_{n \to \infty} n = n \).

\[
\text{Definition 6. If} \ p_n \text{ is a sequence, we define the limit superior of} \ \{p_n\} \text{ as} \ \lim_{n \to \infty} \sup_{k \geq n} p_k \text{ exists.} \]

We call it \(\lim_{n \to \infty} p_n\).
Suppose that the joint pdf involves an unknown parameter \( \theta \). Let’s maximize it:

\[ \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2} \sum_{i=1}^{n} x_i} = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2} \sum_{i=1}^{n} (\theta x_i - \frac{x_i^2}{2})} \]

to get as close as possible to \( \theta \). Assume \( \Theta = (-\infty, \infty) \). We look at \( f_{x_1,\ldots,x_n}(x_1,\ldots,x_n) \) and maximize it over \( \theta \).

Let’s maximize it:

\[ (2\pi)^{n/2} \frac{d}{d\theta} f_{x_1,\ldots,x_n}(x_1,\ldots,x_n) = -\theta \sum_{i=1}^{n} x_i e^{-\frac{1}{2} \sum_{i=1}^{n} x_i} + e^{-\frac{1}{2} \sum_{i=1}^{n} x_i} \]

\[ = 0 \].

Conclusion: \( \hat{\theta}_n \) converges in probability to 0.

14.3. Weak Law of Large Numbers. If \( x_1, x_2, x_3, \ldots, x_n \) are iid and \( E|x_1| < \infty \). Assume \( \text{var}(x_1) < \infty \). Then if we form \( z_n = \frac{1}{n} \sum_{i=1}^{n} x_i \), then \( z_n \) converges to \( E[x_1] \) in probability.

Proof: 

\[ P(|z_n - E[z_n]| > \epsilon) = P(|z_n - E[x_1]| > \epsilon) \leq \frac{\text{var}(z_n)}{\epsilon^2} = \frac{\text{var}(x_1)}{n \epsilon^2} \rightarrow 0. \]

Moreover: If \( E[x_1] = \infty \), then we can prove that \( z_n \rightarrow \infty \) in probability. \( \lim_{n \rightarrow \infty} P(z_n > m) = 1 \).

14.4. Maximum Likelihood Estimation. \( x_1, x_2, \ldots, x_n \). Suppose we have a joint pdf: \( f_{x_1,\ldots,x_n}(x_1,\ldots,x_n) \). Suppose that the joint pdf involves an unknown parameter \( \theta \in \Theta \). Given that the data \( x_1,\ldots,x_n \) is available, we look at \( f_{x_1,\ldots,x_n}(x_1,\ldots,x_n) \) and find \( \theta_0 \in \Theta \) that maximizes it. We call such \( \theta_0 \) the maximum likelihood estimate of \( \theta \).

**Example 20.** \( x_1, \ldots, x_n, x_1 \sim N(\theta, 1) \) and are iid. \( f_{x_1,\ldots,x_n}(x_1,\ldots,x_n) = \prod_{i=1}^{n} f_{x_i}(x_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i-\theta)^2}{2}} = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2} \sum_{i=1}^{n} (\theta x_i - \frac{x_i^2}{2})} \). Assume \( \Theta = (-\infty, \infty) \). We look at \( f_{x_1,\ldots,x_n}(x_1,\ldots,x_n) \) and maximize it over \( \theta \).

Let’s maximize it:

\[ (2\pi)^{n/2} \frac{d}{d\theta} f_{x_1,\ldots,x_n}(x_1,\ldots,x_n) = -\theta \sum_{i=1}^{n} x_i e^{-\frac{1}{2} \sum_{i=1}^{n} x_i} + e^{-\frac{1}{2} \sum_{i=1}^{n} x_i} \sum_{i=1}^{n} x_i \]

\[ = 0 \].

\( \theta_0 = \frac{1}{n} \sum_{i=1}^{n} x_i \).

**Example 21.** Suppose that \( x_i \) has the following pdf:

\[
\begin{array}{c|c}
\hline
x_i & f_{x_i}(x) \\
\hline
a & \frac{1}{a} \\
\hline
\end{array}
\]

\( a \) is unknown. \( \hat{a}_{M LE} = \max(x_1 \ldots x_n) \) to get as close as possible to \( a \).
15. Lecture 13: 10/15/03

15.1. Maximum Likelihood Estimation. If \( x_1 \sim N(\theta, 1) \), then \( \hat{\theta}_{ML} = \frac{1}{n} \sum_{i=1}^{n} x_i \). Suppose \( x_1, \ldots, x_n \) are iid and

\[
 f_{x_1}(x) = \begin{cases} 
 \frac{1}{a}, & x \in [0, a] \\
 0, & \text{else}
\end{cases}
\]

(14)

What is \( \hat{\alpha}_{ML} \)? \( f_{x_1,\ldots,x_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{x_i}(x_i) = \prod_{i=1}^{n} \frac{1}{a} I_{[0,a]}(x_i) = \frac{1}{a^n} \prod_{i=1}^{n} I_{[0,a]}(x_i) \). We want \( a \) to be just larger than all the \( x \)'s: Set \( \hat{\alpha}_{ML} = \max(x_1, \ldots, x_2) \).

Example 22.

\[
 f_{x_i}(x) = \begin{cases} 
 \frac{1}{b-a}, & x \in [a, b] \\
 0, & \text{else}
\end{cases}
\]

(15)

\( \hat{\alpha}_{ML} = \min(x_1, \ldots, x_n) \) and \( \hat{\beta}_{ML} = \max(x_1, \ldots, x_n) \).

Consider: \( \hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} x_i \). This is a linear estimator of \( \theta \) since \( \hat{\theta} \) is a linear function of the data. In contrast, \( \hat{\alpha}_{ML} \) is not a linear estimator. \( x_i \sim f(x) \) parameterized by \( \theta \in \Theta \). We know \( E[x_i] = \theta \).

Question: (\( \hat{\theta} \)) Is this necessarily the minimum square error estimator? That is: Does \( \hat{\theta} \) minimize \( E[(\hat{\theta} - \theta)^2] \) for any \( \hat{\theta} \), another estimator?

Answer: No. Take \( \tilde{\theta} = a \hat{\theta} \), where \( "a" \) is some parameter \( \to \hat{\theta}^* = \frac{a}{n} \sum_{i=1}^{n} x_i \) and let us compute the mean square error: \( E[(\hat{\theta}^* - \theta)^2] = E[(\hat{\theta}^* - \tilde{\theta} + \tilde{\theta} - \theta)^2] \)
\[
\tilde{\theta}^* = E[\hat{\theta}^*] = a\theta = E[(\hat{\theta}^* - \tilde{\theta})^2] + E[(\tilde{\theta} - \theta)^2] + 2E[(\hat{\theta}^* - \tilde{\theta})(\tilde{\theta} - \theta)].
\]

After minor manipulation: \( E[(\hat{\theta}^* - \theta)^2] = \text{var}(\hat{\theta}^*) + (\hat{\theta}^* - \theta)^2 = \text{variance of } \hat{\theta}^* + (\text{bias})^2 \). This is always true. There is an uncertainty principle governing bias and variance of an estimator. \( E[(\hat{\theta}^* - \theta)^2] = \frac{1}{2a^2} \sigma^2_x + (a - 1)^2 \theta^2 \), where \( \sigma^2_x = \text{var}(x_i) \). Now we observe that MSE depends on \( n \) and \( \theta \), so we can try to pick \( a \) that minimizes the MSE. In particular, if \( a = a^* = \frac{n}{2\sigma^2_x} \), then the MSE is at a minimum.

Conclusion: \( \hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} x_i \) is NOT the minimum MSE estimator.

Observation: In route to finding a smaller error, we came up with a formula for \( a^* \) that depends on \( \theta \). \( \therefore \) it will not help us.

15.2. Linear Parameter Estimation. Put \( \hat{\theta} = a^T \mathbf{x} \), where

\[
a = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}
\]

(16)
Problem: Find $a$ such that $\text{MSE} = E[(\hat{\theta} - \theta)^2]$ is minimized. Solution: $\text{MSE} = E[a^T x - \theta]^2 = E[a^T x a - 2\theta a^T x + \theta^2]$. \(\forall a\) MSE $= 0 \rightarrow a_i = \frac{E[x_i]}{\theta^2}$; still depends on $\theta$.

Alternative problem: Suppose that we observe the parameter $\theta$ plus some noise (or uncertainty).

Consider the following model: $Y = \theta^T + N$, where

\begin{equation}
Y = \begin{bmatrix}
Y_1 \\
\vdots \\
Y_n
\end{bmatrix}
\end{equation}

is the observation vector.

\begin{equation}
H = \begin{bmatrix}
h_1 \\
\vdots \\
h_n
\end{bmatrix} \text{ (known)} \quad N = \begin{bmatrix}
N_1 \\
\vdots \\
N_n
\end{bmatrix} \text{ (noise)}.
\end{equation}

For example,

\begin{equation}
H = \begin{bmatrix}
1 \\
\vdots \\
1
\end{bmatrix}
\end{equation}

and $Y_i = \theta + N_i$. Find a linear (in $Y$) estimator $\hat{\theta}$ which minimizes the MSE $= E[(\hat{\theta} - \theta)^2]$. Namely, $\hat{\theta} = a^T Y$. It turns out that $a = \theta^2 (C_{YY})^{-1} H$, where $C_{YY} = E[YY^T]$. $C_{YY} = E[(\theta H + N)(\theta H + N)^T] = \theta^2 HH^T + C_{NN}$, assuming $E[N] = 0$. Also, if

\begin{equation}
H = \begin{bmatrix}
1 \\
\vdots \\
1
\end{bmatrix},
\end{equation}

then $\theta^2 I$ since $HH^T = I$. $C_{YY} = \theta^2 I + C_{NN}$. The noise components are independent. If $C_{NN} = \sigma_N^2 I$, then $C_{YY} = (\theta^2 + \sigma_N^2) I$. Then, $C_{YY}^{-1} = \frac{1}{\theta^2 + \sigma_N^2} I$.

\begin{equation}
a = \frac{\theta^2 H}{\sigma_N^2 + \theta^2} = \frac{\theta^2}{\sigma_N^2 + \theta^2} \begin{bmatrix}
1 \\
\vdots \\
1
\end{bmatrix},
\end{equation}

which depends on $\theta$. If $\sigma_N^2$ is negligible relative to $\theta$, then

\begin{equation}
a = \begin{bmatrix}
1 \\
\vdots \\
1
\end{bmatrix}.
\end{equation}

Look back at the case: $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n x_i$ for $\hat{\theta}^*$, $a^* = \frac{n}{\sigma^2 + n}$. Again, if $\sigma^2 \ll \theta^2$, $a^* \sim 1$. 

The last observation model makes sense: Assume $\theta = E[x_i]$. $x_i = \theta + x_i - \theta = \theta + (x_i - \theta)$ → assumption of zero-mean noise makes sense.

**Alternative approach:** (also a MSE-based approach). Consider minimizing the following modified MSE: $e^2 = E[(Y - \hat{H}\hat{\theta})^2]$ over $\hat{\theta}$.

Solution: $\hat{\theta} = (H^TH)^{-1}H^TY$. The optimal estimate (in this sense) is a linear estimator that does not depend on $\theta$ anymore. If

$$
H = a = \frac{\theta^2H}{\sigma_Y^2 + \theta^2} = \frac{\theta^2}{\sigma_Y^2 + \theta^2} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix},
$$

then $(H^TH)^{-1}H^T = (\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n})$. 

16. Lecture 14: 10/20/03

16.1. Example of MLE. \( x_1, \ldots, x_n \) iid. \( x_1 \in \{0, 1\} \). \( P(x = 1) = p \). \( p \) is our unknown parameter. What is the MLE of \( p \) for discrete-valued data? In general, \( \hat{\theta}_{\text{MLE}} \) maximizes over all \( \theta \in \Theta \), the joint Probability Mass Function (PMF) \( f_{X_1, \ldots, X_n}(a_1, \ldots, a_n) = p\{x_1 = a_1, x_2 = a_2, \ldots, x_n = a_n\} \). In our example, \( f_{X_1, \ldots, X_n}(a_1, \ldots, a_i, \ldots, a_j) \). If we see \( a_i \), \( f_{X_i}(a_i) = p \hat{\theta}_1(a_i) \).

\[
\begin{array}{c|c|c}
0 & \cdot & 1 \\
\hline
1-p & \cdot & a_i
\end{array}
\]

**Figure 29.**

\[
f_{X_1}(a_1) \ldots f_{X_n}(a_n) = p^{\sum_{i=1}^{n}=a_i} (1-p)^{n-\sum_{i=1}^{n}=a_i}. \frac{d}{dp} f_{X_1}(a_1) \ldots f_{X_n}(a_n) = 0. \] Find the corresponding maximizing \( p \in [0, 1] \). We call it \( \hat{p}_{\text{ML}} = \frac{1}{n} \sum_{i=1}^{n}=a_i \).

16.2. Almost Sure Convergence. Suppose that \( \{x_i\}_{i=1}^{\infty} \) is a sequence of r.v.’s. We say \( x_i \) converges to a r.v. \( x \) almost surely (a.s.) if \( P(\lim_{n \to \infty} x_n = x) = 1 \). Also called convergence with probability 1. Recall that convergence in probability is \( \lim_{n \to \infty} P(|x_n - x| > \epsilon) = 0, \forall \epsilon > 0. \)

**Example 23.** Strong Law of Large Numbers (Kolmogorov). If \( x_1, x_2, \ldots \) are iid.

**Example 24.** If \( E[|x_1|] < \infty \), then \( \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n}=x_i = E[x_1] \) a.s. (see proof in text). Moreover, if \( E[x_1] = +\infty \), then \( \frac{1}{n} \sum_{i=1}^{n}=x_i \to +\infty \) (e.g., Cauchy r.v.).

16.3. Another Type of Convergence of Random Sequences. We say \( x_n \) converges to r.v. \( x \) in distribution (weakly) if \( \lim_{x \to y} F_{X_n}(x) = F_X(y) \) at every continuity point \( y \) of \( F_X(y) \) (write \( x_n \to x \)).

Consequence: If \( x_n \) and \( x \) have pdf’s (for every \( n \)) then if \( x_n \to x \), then \( \lim_{n \to \infty} f_{X_n}(x) = f_X(x) \) almost everywhere (length of the subset of \( \mathbb{R} \) over which convergence does not occur is zero).

16.4. Central Limit Theorem (CLT). Let \( x_1, x_2, \ldots \) be iid. \( \text{var}(x_1) = \sigma_x^2 < \infty \) (automatically \( E[|x_1|] < \infty \)). Assume \( E[x_1] = 0 \) and \( \sigma_x^2 = 1 \). Let \( z_n = \sum_{i=1}^{n} \frac{x_i}{\sqrt{n}} \). Then, \( z_n \) converges to \( x \) where \( x \sim N(0, 1) \) (\( z_n \to x \)).

Proof: We need the concept of a characteristic function (cf). Let \( Y \) be a r.v. We define the \( \text{cf} \) of \( Y \) as: \( \phi_Y(\omega) = E[e^{iuY}] \), where \( E[e^{iuY}] \) exists if \( E[\text{Re}(e^{iuY})] < \infty \) and \( E[\text{Im}(e^{iuY})] < \infty \). \( \text{Re}(e^{iuY}) \leq |e^{iuY}| = 1 \). \( \text{Im}(e^{iuY}) \leq 1 \). \( E[|\text{Re}(e^{iuY})|] \leq E[1] = 1 \). \( E[|\text{Im}(e^{iuY})|] \leq 1 \). \( E[e^{iuY}] \) is well defined, \( u \in \mathbb{R} \). If \( Y \) has a pdf \( f_Y(y) \), then \( E[e^{iuY}] = \int_{-\infty}^{\infty} e^{iuY} f_Y(y)dy \).
Example 25. \( Y \in \{0, 1\} \) and \( p\{Y = 1\} = p \). \( \phi_Y(u) = E[e^{iuY}] = pe^{iu} + (1 - p)e^{i0} = pe^{iu} + (1 - p) \).

Example 26. \( Y \sim N(0, 1) \). Then, \( E[e^{iuY}] = e^{-\frac{u^2}{2}} \) (see text).

16.5. Inversion Lemma (Levy’s).

Lemma 1. If \( \phi_x(n) \) is the cf of \( x \), then \( \lim_{c \to \infty} \frac{1}{2\pi} \int_{-c}^{c} e^{iuX - e^{iub}} \phi_x(u) du = P\{a < x < b\} + \frac{P(x=a) + P(x=b)}{2} \) [1].

Fact: If \( \phi_x(u) = \phi_Y(u) \), then \( P\{X = Y\} = 1. P\{X \neq Y\} = 0. \) : the cf fully characterizes a r.v.

Consequence: \( \phi_Y(u) = e^{-\frac{u^2}{2}} \). Then, \( Y \sim N(0, 1) \) with probability 1.

If \( x_1 \ldots x_n \) is a sequence, then we write \( \phi_{x_1 \ldots x_n}(u_1, \ldots, u_n) = E[e^{iu_1x_1 + iu_2x_2 + \ldots + iu_nx_n}] = \) the joint cf function of the vector \( x = (x_1 \ldots x_n) \).

If the \( x_i \)'s are independent: \( \phi_{x_1 \ldots x_n}(u_1, \ldots, u_n) = E[\prod_{j=1}^{n} e^{iu_jx_j}] = \prod_{j=1}^{n} E[e^{iu_jx_j}] = \prod_{j=1}^{n} \phi_{x_j}(u_j) \).

Fact: If \( X \) and \( Y \) are independent, then \( E[g(x)h(y)] = E[g(x)]E[h(y)] \).

Sums of independent r.v.'s: Suppose that \( z = x_1 + x_2 + \ldots + x_n \) and the \( x_i \)'s are iid. Then, \( \phi_{z}(n) = E[e^{iuZ}] = E[e^{iu(x_1 + x_2 + \ldots + x_n)}] = \phi_{x_1}(u)^n = (\text{F.T. of } f_{x_1}(x) \ast \ldots \ast f_{x_1}(x) - n \text{ times}) \).

16.6. Back to Proof of CLT. \( z_n = \frac{x_1 + x_2 + \ldots + x_n}{\sqrt{n}} = \sum_{j=1}^{n} \frac{x_j}{\sqrt{n}} \) and \( x_j \)'s are iid. \( \phi_{x_j}(u) = \left[ \phi_{\frac{x_j}{\sqrt{n}}}(u) \right]^n = [E[e^{iu\frac{x_j}{\sqrt{n}}}]^n = [E[e^{iu\frac{X}{\sqrt{n}}}]^n = \phi_{x_1}(\frac{u}{\sqrt{n}})^n \). Next step: Show \( \lim_{n \to \infty} \phi_{x_1}(\frac{u}{\sqrt{n}})^n = e^{-\frac{u^2}{2}} \).
17. Lecture 15: 10/22/03

17.1. Central Limit Theorem (CLT). Suppose that \(x_1, x_2, \ldots\) are iid. \(E[x_1] = 0, E[x_1^2] = \sigma^2 < \infty\). Then \(z_n = \sum_{i=1}^{n} \frac{x_i}{\sqrt{n}}\). Then, as \(n \to \infty\), \(z_n\) converges to a zero-mean unit-variance Gaussian r.v. in distribution \((z_n \Rightarrow z), z \sim N(0,1)\).

Proof: Let \(\phi_x(u)\) be the cf of \(x_1\). Let \(W = \frac{1}{\sqrt{n}}\phi_x(u) = \phi_x\left(\frac{u}{\sqrt{n}}\right)\). Moreover, \(\phi_{Z_n}(u) = \phi_W(u)^n = \phi_x\left(\frac{u}{\sqrt{n}}\right)^n\). We need to show that: \(\lim_{n \to \infty} \phi_{Z_n}(u) = e^{-\frac{u^2}{2}}\). Equivalently, we need to show: \(\lim_{n \to \infty} \log \phi_{Z_n}(u) = -\frac{u^2}{2}\) or \(\lim_{n \to \infty} n \log \phi_W(u) = -\frac{u^2}{2}\). \(\phi_W(u) = \sum_{i=0}^{\infty} \phi_W^{(i)}(u_0)(u-u_0)^i\), where \(u_0 \leq u = \sum_{i=0}^{L-1} \phi_W^{(l)}(u_0)\left(\frac{u-u_0}{n}\right)^i + \phi_W^{(L)}(\zeta)\left(\frac{u-u_0}{n}\right)^L\); where \(\zeta \in [u_0, u]\).

17.2. Taylor’s Theorem with a Remainder. Pick \(L = 3\). \(\phi_W(u) = \phi_W(0) + \phi_W^{(1)}(0)(u) + \phi_W^{(2)}(0)(u^2) + \phi_W^{(3)}(0)(u^3)\). \(\phi_W(0) = E[e^{iuw}] = 1\). \(\phi_W^{(1)}(0) = \frac{d}{du}E[e^{iuw}]|_{u=0} = \frac{d}{du}E[e^{iuw}]|_{u=0} = E \left[ \frac{d}{du} e^{iuw} \right] |_{u=0} = E[i \frac{d}{du} e^{iuw}] |_{u=0} = 0\). \(\phi_W^{(2)}(0) = \ldots = -\frac{1}{n^2}\) (assume \(\sigma^2 = 1\)). \(\phi_W^{(3)}(\zeta)\) = \(\frac{d^3}{du^3} \int_{-\infty}^{\infty} e^{iuw} f_w(w) dw |_{u=\zeta}\). \(f_W(w) = f_z(w \sqrt{n})\sqrt{n}\). \(\phi_W^{(3)}(\zeta) = \frac{d^3}{du^3} \int_{-\infty}^{\infty} e^{iuw} f_z(w \sqrt{n}) dw \sqrt{n} = \int_{-\infty}^{\infty} (iw)^3 e^{iuw} f_z(w \sqrt{n}) dw = \frac{1}{n \sqrt{n}} R(n, \zeta) \Rightarrow \phi_W(u) = 1 - \frac{1}{2n^2} u^2 + \frac{1}{n \sqrt{n}} R(n, \zeta), \) for some \(\zeta\). \(n \log \phi_W(u) = n \log(1 - \frac{1}{2n^2} u^2 + \frac{1}{n \sqrt{n}} R(n, \zeta))\).

\[ \begin{align*}
\text{Figure 30.} & \quad \text{The log function can be approximated by a 45° line.} \\
& \quad \text{For any } u, \text{ we can make } n \text{ large enough so that } -\frac{1}{2n^2} u^2 + \frac{1}{n \sqrt{n}} R \text{ is smaller than 1. Note that for} \\
& \quad x < 1, \log(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots, \text{ and } n \log \phi_W(u) = n \left\{ \left( -\frac{1}{2n^2} u^2 + \frac{1}{n \sqrt{n}} R \right) - \frac{x^3}{3} + \frac{x^4}{4} + \ldots \right\} \\
& \quad = -\frac{1}{2n^2} u^2 + \frac{1}{n \sqrt{n}} R - \left( -\frac{1}{2n^2} + \frac{1}{n} \right) R^2/2! + \ldots \text{ limits the } n \log \phi_W(u) = -\frac{1}{2n^2} u^2. \)
\end{align*} \]

Simple Extension: If \(x_1, x_2, \ldots\) iid, \(E[x_1] = a\), and \(\text{var}(x_1) = \sigma^2\), then \(z_n \Rightarrow \frac{\sum_{i=1}^{n} (x_i - a_i)}{\sqrt{n} \sqrt{\sum_{i=1}^{n} \sigma_i^2}}\). \(z_n \Rightarrow z\), where \(z\) is a zero-mean, unit-variance Gaussian r.v.

Another Extension [Lindeberg]: If \(x_1, x_2, \ldots\) are independent, \(E[x_1] = a_i < \infty\), \(\text{var}(x_i) = \sigma_i^2\) and \(\sum_{i=1}^{n} \sigma_i^2 < \infty\). Plus, another technical condition (Lindeberg condition). Then, \(z_n \Rightarrow \frac{\sum_{i=1}^{n} (x_i - a_i)}{\sqrt{n} \sqrt{\sum_{i=1}^{n} \sigma_i^2}}\).

Fact: If \(x_n \to x\) a.s., then \(x_n\) converges to \(x\) in probability.
17.3. **Whitening Filter.** Preliminaries on covariance matrices. If \( \mathbf{x} = [x_1 \ldots x_n]^T \), then the covariance matrix is: \( \mathbf{C}_X \) is a \( n \times n \) matrix = \( E[(\mathbf{x} - E[\mathbf{x}]) (\mathbf{x} - E[\mathbf{x}])^T] \).

Properties:

1. \( \mathbf{C}_X \) is symmetric.
2. There is a transformation \( \mathbf{A} \) such that \( \mathbf{Y} = \mathbf{A} \mathbf{X} \) has a diagonal covariance matrix.
18. Lecture 16: 10/27/03

18.1. **Covariance Matrix.** Let \( \mathbf{x} \) be a random vector with covariance matrix \( \mathbf{C}_\mathbf{x} = E[(\mathbf{x} - E[\mathbf{x}])(\mathbf{x} - E[\mathbf{x}])^T] \). \( \mathbf{C}_\mathbf{x} \) has the following properties:

(1) \( \mathbf{C}_\mathbf{x} \) is symmetric.

(2) There is a matrix \( \phi \) such that \( \phi^T \mathbf{C}_\mathbf{x} \phi = \Lambda \) is diagonal. In particular, if \( \lambda_1 \ldots \lambda_n \) are the eigenvalues of \( \mathbf{C}_\mathbf{x} \) (assume distinct) with corresponding eigenvectors \( \mathbf{v}_1 \ldots \mathbf{v}_n \), then \( \phi = [\mathbf{v}_1; \mathbf{v}_2; \ldots; \mathbf{v}_n] \). Moreover, \( \phi \) can be selected so that the eigenvectors are orthogonal. That is,

\[
\mathbf{v}_i^T \mathbf{v}_j = \begin{cases} 
1, & i = j \\
0, & i \neq j
\end{cases}
\]

If \( \mathbf{C}_\mathbf{x} \) is symmetric then eigenvectors corresponding to distinct eigenvalues are orthogonal.

(3) Since the \( \mathbf{v}_i \)'s are orthogonal, they span \( \mathbb{R}^n \). So, we obtain the following representation.

\[
\mathbf{x} = \sum_{i=1}^n c_i \mathbf{v}_i \text{ for any } \mathbf{x} \in \mathbb{R}^n \text{ and } c_i = \mathbf{x}^T \mathbf{v}_i.
\]

(4) \( \mathbf{C}_x \) is positive semi-definite, meaning: \( \mathbf{x}^T \mathbf{C}_\mathbf{x} \mathbf{x} \geq 0 \).

Proof: \( \mathbf{x}^T \mathbf{C}_\mathbf{x} \mathbf{x} = \mathbf{x}^T E[(\mathbf{x} - E[\mathbf{x}])(\mathbf{x} - E[\mathbf{x}])^T] \mathbf{x} = E[(\mathbf{x}^T(\mathbf{x} - E[\mathbf{x}]))^2] \geq 0 \).

(5) Unless one or more components of \( \mathbf{x} \) is an affine superposition of the other components (e.g., \( x_1 = 2x_2 + 0.5x_3 + 7 \)), \( \mathbf{C}_\mathbf{x} \) is positive definitive.

Proof: Note that \( E[Y^2] = 0 \) iff \( Y = 0 \) almost surely (\( P\{Y = 0\} = 1 \)). In the homework, you looked at an estimate of \( \mathbf{C}_\mathbf{x} \), defined by \( \hat{\mathbf{C}}_\mathbf{x} = \mathbf{x}\mathbf{x}^T \). \( \mathbf{C} = \frac{1}{K} \sum_{i=1}^K \mathbf{x}^{(i)} \mathbf{x}^{(i)T} \) (assume \( E[\mathbf{x}] = 0 \)), which has rank at most 1. \( \ldots \) if \( n \geq 2 \), \( \mathbf{C}_\mathbf{x} \) is not full rank. \( \ldots \) \( \hat{\mathbf{C}}_\mathbf{x} \) is not positive definite.

18.2. **Whitening of Correlated Data.** Let \( x_1 \ldots x_2 \) be a sequence. We say that the components are white if: \( E[(x_i - E[x_i])(x_j - E[x_j])] = 0 \), \( \forall i \neq j \).

Let \( x_1 \ldots x_n \) be given, call it

\[
x = \begin{bmatrix} x_1 \\
\vdots \\
x_n \end{bmatrix},
\]

with \( \mathbf{C}_x \). Then, there is a matrix \( \mathbf{A} \) such that if \( \mathbf{Y} = \mathbf{A}\mathbf{X} \), then \( \mathbf{C}_\mathbf{Y} = \mathbf{I} \).

Proof: Put \( \mathbf{B} = \phi \mathbf{A}^{1/2} \), where

\[
\Lambda = \begin{bmatrix} 
\lambda_1 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \lambda_n
\end{bmatrix},
\]

and \( \lambda_i \) is the \( i \)th eigenvalue of \( \mathbf{C}_x \).
Back to Property 2, \( \phi^T C \phi = \Lambda \). What is \( B^{-1} \phi^T \)? \( B(\Lambda^{-1/2} \phi^T) = \phi \Lambda^{-1/2} \Lambda^{-1/2} \phi^T = \phi I \phi^T \)

\( = \phi \phi^T = I \). \( B^{-1} = \Lambda^{-1/2} \phi^T \). Take \( A = B^{-1} \). Claim: If \( Y = AX \), then \( C_Y = AC_X A^T = \Lambda^{-1/2} \phi^T C \phi \Lambda^{-1/2} = \Lambda^{-1/2} \Lambda \Lambda^{-1/2} = I \).

Filter? \( A = \Lambda^{-1/2} \phi^T \). \( AX = \Lambda^{-1/2} \phi^T X \) =

\[
\begin{pmatrix}
\ddots & \cdots & \cdots & \cdots \\
& & & \\
& & & \\
& & & \\
& & & \\
& & & \\
& & & \\
& & & \\
\end{pmatrix}
\]

(27)

If \( Y_k \) is the \( i \)th con

with \( h_{ij} = \frac{1}{\sqrt{\lambda_i}}(v_i) \).

\[
\begin{array}{c}
X_K \\
\text{Whitening} \\
\text{Filter \{h\}} \\
\rightarrow \\
\end{array}
\]

\[
Y_K
\]

white sequence

Figure 31.

Recall the representation \( x = \sum_{k=1}^{n}(v_k^T x)v_k \). Also, \( Y_k = \frac{1}{\sqrt{\lambda_k}} Y_k v_k \), which is known as the reconstruction formula.

18.3. **Simultaneous Diagonalization of Two Covariance Matrices.** Assume everything as before. In addition, suppose that \( C_Z \) is a second covariance matrix (corresponding to a random vector \( Z \)). Goal: Find a transformation \( \Gamma \) so that both of \( C_X \) and \( C_Z \) are diagonalizable.

I would try \( A \) again: We know that \( A \) whitens \( X \). Put \( T = AZ \). \( C_T = AC_Z A^T = \Lambda^{-1/2} \phi^T C \phi \Lambda^{-1/2} \) = ? Note that \( W \) is eigenvector of \( C_T \). We can diagonalize \( G \) as well. Let \( \eta_1 \ldots \eta_n \) be the eigenvectors of \( G \). Then we can form a matrix \( W \) (like \( \phi \)) such that \( W^T GW = M \), where

\[
M = \begin{bmatrix}
\eta_1 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \eta_n
\end{bmatrix}
\]

Let’s try \( \Gamma = W^T A \). Check: (1) Consider \( D = \Gamma X \). \( C_D = W^T AC_X A^T W = W^T IW = W^T W = I \) (because \( W \) is normal matrix, i.e., \( W^{-1} = W^T \)). (2) \( F = \Gamma Z \). \( C_F = \Gamma C_Z \Gamma^T = W^T A C_Z A^T W \). But, \( AC_Z A^T = G \Rightarrow C_F = W^T GW = M \).

18.4. **Summary.** Let \( C_X \) and \( C_Z \) be given. Let \( B \) be as before and let \( W \) be the matrix whose \( i \)th column is the eigenvector of \( B^{-1} C_Z B \) corresponding to the \( i \)th eigenvalue. Form \( \Gamma \) as \( W^T B^{-1} \).

Then \( Y = \Gamma X \) and \( T = \Gamma Z \) have the following properties:

(1) \( C_Y = I \).
(2) \( C_T = M = \)
\[
\begin{bmatrix}
\eta_1 & \ldots & 0 \\
\vdots & \ddots & \vdots \\
0 & \ldots & \eta_n
\end{bmatrix}
\]

(29) where the \( \eta_i \)'s are the eigenvalues of \( B^{-1}C_z B \).

Example 27.

(30) \( C_X = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} ; C_Z = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix} \).

(1) Find \( \phi \): eigenvalues: \( |C_X - \lambda I| = 0 \). \( \lambda_1 = 1, \lambda_2 = 3 \). Find \( v_1 \) and \( v_2 \). \( C_X v_1 = \lambda_1 v_1 \) and \( C_X v_2 = \lambda_2 v_2 \). Normalize so that \( ||v_1|| = ||v_2|| = 1 \).

(31) \( v_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} ; v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \).

(2)

(32) \( B^{-1} = A^{-1/2} \phi^T = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \).

(3)

(33) \( B^{-1} G (B^{-1})^T = \frac{1}{2} \begin{bmatrix} 8 & 0 \\ 0 & 4 \end{bmatrix} \).

\( \eta_1 = 4 \) and \( \eta_2 = \frac{2}{3} \). \( \therefore W = I \).

(34) \( \Gamma = W^T B^{-1} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{bmatrix} \).
19. Lecture 17: 10/29/03

19.1. Random Processes. Let \((\Omega, \mathcal{F}, P)\) be a probability space. Then define a mapping \(X\) from the sample space \(\Omega\) to a space of continuous functions. Each element of this space is called a sampling function. This mapping is called a random process if at each fixed time, the mapping is a random variable.

Example 28. Let \(N(t)\) be the number of hits a server receives by time \(t\). \(\Omega\) is the collection of all possible arrival times. Take \(\omega = \{T_1, T_2, T_3, \ldots\}\), \(T_1 \subset T_2 \subset T_3 \subset \ldots\). \(N(t, \omega) = \sum_{i=1}^{\infty} I_{[0,T_i]}(t)\), where

\[
I_{[0,T_i]}(t) = \begin{cases} 
1, & t \in [0,T_i] \\
0, & \text{otherwise} 
\end{cases}
\]

\(N(t, \omega) = \sum_{i=1}^{\infty} u(t - T_i).\)

\(\omega\) represents all arrival times in a given period. So, a process can be thought of as a function of two variables, time and r.v. (arrival time). For each \(\omega\), \(N(t, \omega)\) is a function of \(t\). So, we think of a realization of a random process as a function of time instead of a scalar number in the case of a random variable.

Example 29. Let \(A\) be a \([-1,1]\)-valued binary r.v. Then, define \(x(t) = A \cos(2\pi f_0 t)\), where \(P\{A = 1\} = P\{A = -1\} = 1/2\). \(\Omega = \{-1, 1\}\) for \(t = t_1\), \(x(t_1) = A \cos(2\pi f_0 t_1)\): scaled r.v. So, \(x(t_1) \equiv x_1: \text{r.v. evaluated at time } t_1\), so we can talk about its pdf. pdf: \(f_{x(t_1)}(x)\) or \(f_x(x_1, t_1)\) or \(f_{x_1}(x_1)\).

Assume I have \(x(t_2) \equiv x_2\). joint pdf: \(f_{x_1(t_1), x_2(t_2)} = f_{x_1, x_2}(x_1, x_2) = f_x(x_1, x_2; t_1, t_2)\), and so on.

Mean: \(E[x(t_1)] = \int_{-\infty}^{\infty} x(t_1, a) f_A(a) da = \int_{-\infty}^{\infty} x_1 f_{x_1}(x_1) dx_1.\)

Correlation:

\[
R_x(t_1, t_2) = E[x(t_1)x(t_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f_{x_1, x_2}(x_1, x_2) dx_1 dx_2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f_x(x_1, x_2; t_1, t_2) dx_1 dx_2.
\]

Covariance: \(C_x(t_1, t_2) = E[(x(t_1) - \mu_x(t_1))(x(t_2) - \mu_x(t_2))]\), \(\mu_x(t) = E[x(t)].\)
19.2. **Wide Sense Stationary (WSS).** A process is said to be wide sense stationary (WSS) if

1. $R_x(t_1, t_2)$ is constant.
2. $R_x(t_1, t_2) = R_x(t_1 - t_2)$.

$F_{x_1, x_2, \ldots, x_n}(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n) = F_{x_1, x_2, \ldots, x_n}(x_1, x_2, \ldots, x_n; t_1 + T, t_2 + T, \ldots, t_n + T)$. If this is true for any $t_1, t_2, \ldots, t_n$ and for any $n$, then $x(t)$ is a stationary random process. For WSS, $F_{x_1}(x_1, t) = F_{x_1}(x_1, t_1 + T)$. $F_{x_1, x_2}(x_1, x_2; t_1 + T, t_2 + T) = F_{x_1, x_2}(x_1, x_2; t_1, t_2)$.

**Example 30.** $x(t) = A \sin(\omega_0 t + \Theta)$, $\Theta$ and $A$ are independent r.v.'s. Let $E[A] = \mu_A$ and assume that $\Theta \sim U[\pi, \pi]$

$\Theta)d\theta = \theta$.

![Figure 33](image)

$R_x(t_1, t_2) = E[x(t_1)x(t_2)] = E[A^2 \sin(\omega_0 t_1 + \Theta) \sin(\omega_0 t_2 + \Theta)] = E[A^2]E[\sin(\omega_0 t_1 + \Theta) \sin(\omega_0 t_2 + \Theta)] = (\sigma_A^2 + \mu_A^2) \frac{1}{2} E[\cos(\omega_0 (t_1 - t_2)) - \cos(2\Theta + \omega_0 t_1 + \omega_0 t_2)] = (\sigma_A^2 + \mu_A^2) \frac{1}{2} \cos(\omega_0 (t_1 - t_2))$. \ldots $x(t)$ is WSS.

19.3. **Example of Random Processes.** Brownian motion (Wiener process): $B(t), t \geq 0$ is called a wiener process if:

1. $B(0) = 0$.
2. For each time $t > 0$, $B(t)$ is Gaussian, zero-mean and has variance $\alpha t$.
3. If $t_1 < t_2 < t_3 < \ldots < t_n$, then $B(t_1), B(t_2)B(t_1), \ldots, B(t_n)B(t_n-1)$ are independent increments.

Construction of $B(t)$ as a random walk: Let $\omega(1), \omega(2), \ldots$ be iid and $\omega(k) = \pm S = \pm 1$ with $P(+1) = P(-1) = 1/2$. Fix $n$ and $T$, where $n$ is the number of walks in a period of time equal to $T$. Define $x_n(k) = \sum_{i=1}^{k} \frac{\omega(i)}{\sqrt{n}} \sqrt{T}$, and $\frac{T}{n}$ is the width of the jump or walk in time $t$ and the size of the jump is $\sqrt{T}$. $k = \frac{T}{\sqrt{n}}$. Let $x_n(t) = x_n([\frac{t}{T}])$. $x_n(0) = 0$. Mean: $E[x_n(k)] = 0$. var($\omega(t)$) = $E[\omega(t)^2] = (+1)^2(1/2) + (-1)^2(1/2) = 1$. var($x_n(k)$) = $\sum_{i=1}^{k} (1)^2 \frac{T}{n} = \frac{kT}{n} \approx t$. Also, let $n \to \infty$ for fixed $t$. $t = \frac{kT}{n}$. \ldots $k$ should go to infinity. So, $\lim_{n \to \infty} \sum_{i=1}^{k} \frac{\omega(i)}{\sqrt{n}} \sqrt{T} = \lim_{k \to \infty} \sum_{i=1}^{k} \frac{\omega(i)}{\sqrt{n}} \sqrt{T}$. $\lim_{k \to \infty} \sum_{i=1}^{k} \frac{\omega(i)}{\sqrt{n}} \sqrt{T} = \lim_{k \to \infty} \sqrt{T} \sum_{i=1}^{k} \frac{\omega(i)}{\sqrt{n} \sqrt{T}} = \lim_{k \to \infty} \sqrt{T} \sum_{i=1}^{k} \frac{\omega(i)}{\sqrt{n} \sqrt{T}}$. From the CLT, as $k \to \infty \sum_{i=1}^{k} \frac{\omega(i)}{\sqrt{n} \sqrt{T}} \sim N(0, 1)$. \ldots $\lim_{n \to \infty} \text{var}(x_n(T)) = t$. 

For WSS, $F_{x_1}(x_1, t) = F_{x_1}(x_1, t_1 + T)$. $F_{x_1, x_2}(x_1, x_2; t_1 + T, t_2 + T) = F_{x_1, x_2}(x_1, x_2; t_1, t_2)$.
\[ B(t) \text{ is not WSS: } R_B(t_1, t_2) = E[B(t_1)B(t_2)]. \text{ Assume } t_1 < t_2. \text{ } R_B(t_1, t_2) = E[B(t_1)(B(t_2) + B(t_1) - B(t_1))] = E[B^2(t_1) + B(t_1)[B(t_2) - B(t_1)]] = E[B^2(t_1)] = t_1. \text{ In general, } R_B(t_1, t_2) = \min(t_1, t_2) \text{ does not depend on the difference.} \]
20. Lecture 18: 11/03/03

Read article by Clark on "Imaging Spectroscopy" [2, 3].

20.1. Review. Recall: \((\Omega, \mathcal{F})\). R.V. \(x\) is \(\mathcal{F}\)-measurable. \(x : \Omega \to \mathbb{R}\). Now, we define a random process: it is a measurable mapping from \(\Omega\) to a space of functions of time. Thus, we write \(X(\omega, t)\) to emphasize the dependence on \(\omega\), but we often drop it and write \(x(t)\).

**Example 31.** \(\Omega = \{H, T\}\).

\[
X(t, \omega) = \begin{cases} 
+\cos t, & \text{if } \omega = H \\
-\cos t, & \text{if } \omega = T
\end{cases}
\]

**Example 32.** Let \(T\) be a fixed number, representing the time horizon. Let \(\omega_1, \omega_2, \omega_3, \ldots\) be iid \([-1, 1]\)-valued iid sequence with \(P(\omega_1 = -1) = 1/2\). Define the random sequence: for a fixed \(n \geq 1\) (\(n\) is the number of small segments we wish to divide \(T\)). We define the random sequence:

\[
X_n(k) = \frac{1}{\sqrt{n}} \sum_{i=1}^{k} w_i \sqrt{T}. \quad E[X_n(k)] = \frac{\sqrt{T}}{\sqrt{n}} \sum_{i=1}^{k} E[\omega_i] = 0. \quad E[X_n^2(k)] = \left(\frac{\sqrt{T}}{\sqrt{n}}\right)^2 \text{var}(\omega_i^2) = \frac{T}{n}k.
\]

Now, we think of the continuous time \(t\), and think of \(k\) as the integer we must multiply \(T/n\) by so that \(kT/n \approx t\). For example, we take \(k = \lfloor \frac{t}{T/n} \rfloor \) where \(\lfloor x \rfloor = \text{the greatest integer less than or equal to } x\). Now, we can define a stochastic process:

\[
X_{n,T}(t) = \begin{cases} 
1/2, & \text{if } n \leq t < \frac{n}{2} \\
1/4, & \text{if } \frac{n}{2} \leq t < \frac{3n}{4} \\
\vdots & \\
\end{cases}
\]

**Figure 34.** As \(n \nearrow \) it reduces the jump size and the curve becomes smoother.

*What is the sample space \(\Omega\)? \(\Omega = \{-1, +1\} \times \{-1, +1\} \times \ldots = \{-1, +1\}^\infty\), which is an infinite Cartesian product (product space), i.e., a member of \(\Omega\) is typically \(\omega = (+1, -1, -1, +1, \ldots)\).*

*What is \(\mathcal{F}\)? Let \(\mathcal{M} = \text{collection of } \omega's \text{ for which only a finite number of components are specified } = \{(\omega_1, \omega_2, \ldots, \omega_i) : i \text{ is finite}\} = \text{collection of all cylinders}. \mathcal{F} = \sigma(\mathcal{M}). \text{ Now we have } (\Omega, \mathcal{F}).X_{n,T}(t) \text{ is a mapping from } \Omega \text{ to the space of piecewise-constant functions}.*

*What are some of the properties of \(X_{n,T}(\cdot)\)? \(E[X_{n,T}(t)] = \frac{\sqrt{T}}{\sqrt{n}} \sum_{i=1}^{\lfloor t/T/n \rfloor} E[\omega_i] = 0. \quad E[X_{n,T}(t)^2] = \left(\frac{\sqrt{T}}{\sqrt{n}}\right)^2 \frac{t}{T/n} \cdot 1. \quad \lim_{n \to \infty} E[X_{n,T}^2] = \lim_{n \to \infty} (\frac{T}{n} T/n) = t.\)
20.2. **Brownian Motion.** \( B(t) \) has the following properties:

1. \( B(0) = 0 \).
2. For each \( t \), \( B(t) \) is a Gaussian r.v. with mean = 0 and variance of \( t \).
3. For \( t_1 < t_2 < t_3 < t_4 \), the r.v.’s \( B(t_2) - B(t_1) \) and \( B(t_4) - B(t_3) \) are independent.

Call \( X_T(t) \) the limit of \( X_{n,T}(t) \) in the mean square sense. \( \lim_{n \to \infty} E[|X_T(t) - X_{n,T}(t)|^2] = 0, \forall t \).

1. Always verified.
2. Mean = 0, var = \( t \). By CLT it is Gaussian.
3. \( \omega \) in each of the ranges are different. Because \( X_T(t_4) - X_T(t_3) \) and \( X_T(t_2) - X_T(t_1) \) depend on exclusively a disjoint set of \( \omega \)'s. \( \therefore X_T(t) \) is a brownian motion.

General fact: If \( X(t) \) is a stochastic process, then for any fixed \( t = t_0 \), \( x(t_0) \) is a r.v.

\[
E[B(t)] = 0, \quad E[B^2(t)] = t. \quad C_B(t_1, t_2) = E[B(t_1)B(t_2)]; \quad \text{for } t_2 \geq t_1 = E[B(t_1)(B(t_2) - B(t_1) + B(t_1))] = E[B(t_1)](B(t_2) - B(t_1)) + B^2(t_1) = E[B(t_1)]E[B(t_2) - B(t_1)] + E[B^2(t_1)] = t_1. \quad C_B(t_1) = \min(t_1, t_2). \quad B(t) \text{ is not WSS since } C_B(1, 3) = 1 \neq 2 = C_B(0, 3 - 1) \neq 0.
\]

20.3. **White Noise.** A stochastic process \( \omega(t) \) is called Gaussian white noise if:

1. \( \omega(t) \) is Gaussian with zero mean and infinite variance (variance very large).
2. \( C_\omega(t_1, t_2) = \delta(t_2 - t_1). \quad \therefore \omega \text{ is WSS.} \)
21. White Noise. Let \( w(t) \) be Gaussian white noise: \( E[w(t)] = 0 \), \( w(t) \) is WSS. \( R_w(t_1, t_2) = R_w(\tau) = \delta(\tau) \). \( w(t) \) is Gaussian for each \( t \) with variance \( \sigma^2 = \infty \). Consider: \( M(t) = \int_0^t w(\tau) d\tau \).

\[
E[M(t)] = E[\int_0^t w(\tau) d\tau] = \int_0^t E[w(\tau)] d\tau = 0.
\]

\[
E[M(t_1)M(t_2)] = E[\int_0^{t_1} w(\tau) d\tau \int_0^{t_2} w(\eta) d\eta] = \int_0^{t_1} \int_0^{t_2} E[w(\tau)w(\eta)] d\eta d\tau.
\]

Assume \( t_1 \leq t_2 \). Since \( \tau < t \), \( E[M(t_1)M(t_2)] = \int_{t_1}^{t_2} \int_{t_1}^{t_2} E[w(\tau)w(\eta)] d\eta d\tau \).

If \( t_1 > t_2 \Rightarrow E[M(t_1)M(t_2)] = t_2 \Rightarrow E[M(t_1)M(t_2)] = \min(t_1, t_2) \). \( M(t) \) is Gaussian because it is a linear “sum” of Gaussian r.v.’s. \( w(0), w(\Delta t), w(2\Delta t), \ldots, w(t) \). And, if we sum a Gaussian r.v., we obtain a r.v.

21.2. Independent Increment Hypothesis. \( M(t_4) - M(t_3) = \int_{t_3}^{t_4} w(\tau) d\tau \). \( M(t_2) - M(t_1) = \int_{t_1}^{t_2} w(\tau) d\tau \). They are independent because they are Gaussian uncorrelated.

Example 33.

\[
f_{XY}(x, y) = \begin{cases} 
\frac{1}{\pi}, & (x, y) \in D \\
0, & \text{elsewhere}
\end{cases}
\]

\( E[XY] = 0, E[X] = 0 \) and \( E 

Now, we conclude that \( M(t) \) satisfies the conditions for a Brownian motion, i.e., the integral of white noise in BM. White noise is the “derivative” of BM.
21.3. Application of White Noise.

\[ R \]

\[ I(t) \text{: random current (noise)} \]

**Figure 37.**

The electrons wander randomly in random directions due to their acquired thermal energy from the lattice. \( I(t) \) is modeled by Gaussian white noise. Also, \( R_I(\tau) = \frac{2K_T T}{R} \delta(\tau) \). \( K_T \) is Boltzman’s constant, \( T \) is temperature in Kelvin and \( R \) is the resistor.

\[ \delta_j(\tau) = \frac{2K_T T}{R} \]

\[ \delta V_j(\tau) = 2K_T TR \]

**Figure 38.**

\[ \text{Norton:} \]

\[ \text{Thevenin:} \]

\[ \text{Figure 39.} \]

21.4. Norton Model for Thermal Noise. \( Y(t) = R(t) \ast I_J(t, \tilde{\omega}), \) where \( \tilde{\omega} \) is the outcome of the experiment, \( = \int_{-\infty}^{\infty} h(t-\tau)E[I_J(\tau)]d\tau \). \( E[Y(t)] = \int_{-\infty}^{\infty} h(t-\tau)E[I_J(\tau)]d\tau = 0 \). If \( x(t) \) is a WSS process with mean \( \bar{x} \), and if \( h \) is a linear time invariant system, then \( E[Y(t)] = \bar{x} \int_{-\infty}^{\infty} h(t)dt < \infty \) if \( h \) is BIBO stable. Moreover, if \( x(t) \) is not WSS with mean \( \bar{x}(t) = E[x(t)], \) then \( E[Y(t)] = h(t) \ast \bar{x}(t) \).
What about the second moment of $Y(t)$? $E[Y^2(t)]^{1/2}$ is the r.m.s. value of noise ($E[Y^2]$ is power in $1\,\Omega$ resistor. It turns out that $E[Y^2(t)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{YY}(\omega) d\omega$, where $S_{YY}(\omega) = |H(\omega)|^2 \cdot S_{XX}(\omega)$ and $S_{XX}(\omega) = \mathcal{F}\{R_X(\tau)\}$, where $S_{XX}(\omega)$ is the power spectral density of the WSS process $x(t)$.

\begin{equation}
X_T(t) = \begin{cases} 
X(t), & |t| < T \\
0, & \text{otherwise}
\end{cases}
\end{equation}

Let’s take the F.T. of $X_T$:

\begin{equation}
\lim_{T \to \infty} \frac{\left|\mathcal{F}\{X_T(t)\}(\omega)\right|^2}{2T} = S_{XX}(\omega) \quad [\text{rad/s}]
\end{equation}
Consider a WSS process $X(t)$. Note that the mean-square of the process $E[X^2(t)]^{1/2}$ represents the rms power in a $1 \Omega$ resistor. What we would like to do is to develop a formula for the power spectral density of $X(t)$. In other words, we want to be able to write:

$$E[X^2(t)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_X(\omega)d\Omega.$$  

$S_X(\omega)$ is the desired power spectrum.

We think of $S_X(\omega)$ as a frequency distribution, or spectrum. This representation is useful because it informs us where (i.e., at which frequencies) the power of $x(t)$ is concentrated. Define:

$$X_T(t) = \begin{cases} X(t), & |t| < T \\ 0, & \text{otherwise} \end{cases}$$

Let $F_{X,T}(t) = \mathcal{F}[X_T(t)]$.

$$F_{X,T}(\omega) = \int_{-\infty}^{\infty} X_T(t)e^{-j\omega t}dt.$$  

From Parseval’s Theorem:

$$\int_{-\infty}^{\infty} X_T^2(t)dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |F_{X,T}(\omega)|^2d\omega$$

$$\int_{-T}^{T} X_T^2(t)dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |F_{X,T}(\omega)|^2d\omega.$$  

$$\frac{1}{2T} \int_{-T}^{T} X_T^2(t)dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|F_{X,T}(\omega)|^2}{2T}d\omega.$$  

Take the expectation of both sides of Eq. (46),

$$\frac{1}{2T} \int_{-\infty}^{\infty} E[X_T^2(t)]dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{E[|F_{X,T}(\omega)|^2]}{2T}d\omega$$

$$E[X_T^2(t)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{E[|F_{X,T}(\omega)|^2]}{2T}d\omega$$

$$\lim_{T \to \infty} E[X_T^2(t)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{E[|F_{X,T}(\omega)|^2]}{2T}d\omega$$

$$E[X_T^2(t)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left( \lim_{T \to \infty} \frac{E[|F_{X,T}(\omega)|^2]}{2T} \right)d\omega$$

$$\therefore S_X(\omega) = \lim_{T \to \infty} \frac{E[|F_{X,T}(\omega)|^2]}{2T}; \text{ valid only for WSS.}$$

Note: This expression is valid for WSS stochastic processes.
22.1. **Practical Interpretation.** Suppose that we make \( n \) measurements of the sample function. In other words, we “observe” the process \( X(t) \) in the interval \([-T, T]\) \( n \) times. At this point, we have \( n \) functions of time. Let us denote them by: \( X_{1,T}(t), X_{2,T}(t), \ldots, X_{n,T}(t) \). To form an estimate of \( S_X(\omega) \) using these observations (sample functions), we replace the expectation by the sample average (mean):

\[
E[|F_{X,T}(\omega)|^2] \approx \frac{1}{n} \{|F_{X_1,T}(\omega)|^2 + \ldots + |F_{X_n,T}(\omega)|^2\},
\]

where \( F_{X_i,T}(\omega) = \mathcal{F}\{X_i(t)\}; \, i \in \{1, 2, \ldots, n\} \). Finally,

\[
S_X(\omega) \approx \frac{1}{T} \sum_{i=1}^{n} |F_{X_i,T}(\omega)|^2.
\]

22.2. **Connection between \( R_X(\tau) \) and \( S_X(\omega) \).**

\[
R_X(\tau) = E[X(t + \tau)x(t)].
\]

**Theorem 8.** (Wiener-Kinichin)

\[
S_X(\omega) = \mathcal{F}\{R_X(\tau)\}.
\]

Conclusion: The power spectral density can be calculated in two ways, depending on what is available:

1. If \( R_X(\tau) \) is known, then \( S_X(\omega) = \mathcal{F}\{R_X(\tau)\} \).
2. If \( R_X(\tau) \) is unknown, then find the estimate of \( S_X(\omega) \) by Eq. (53).

**Proof of Theorem:**

\[
S_X(\omega) = \int_{-\infty}^{\infty} R_X(\tau)e^{-j\omega\tau}d\tau.
\]

\[
S_X(\omega) = \lim_{T \to \infty} \frac{E[|F_{X,T}(\omega)|^2]}{2T}.
\]

\[
|F_{X,T}(\omega)|^2 = F_{X,T}(\omega)F_{X,T}^*(\omega).
\]

\[
= \int_{-T}^{T} e^{-j\omega t_1}dt_1 \cdot \int_{-T}^{T} x^*(t_1)e^{j\omega t_2}dt_2
\]

\[
= \int_{-T}^{T} \int_{-T}^{T} x(t_1)x^*(t_2)e^{-j\omega(t_1-t_2)}dt_1dt_2.
\]

\[
E[|F_{X,T}(\omega)|^2] = \int_{-T}^{T} \int_{-T}^{T} E[x(t_1)x^*(t_2)]e^{-j\omega(t_1-t_2)}dt_1dt_2.
\]

\[
= \int_{-T}^{T} \int_{-T}^{T} R_X(t_1 - t_2).
\]
\( t_1 = \frac{1}{2}s + \frac{1}{2}\tau, \quad t_2 = \frac{1}{2}s - \frac{1}{2}\tau. \quad t_1 = T = \frac{1}{2}s + \frac{1}{2}\tau \rightarrow s = 2T - \tau. \quad t_1 = -T = \frac{1}{2}s + \frac{1}{2}\tau \rightarrow s = -2T - \tau. \)

Get \( t_2 = T \) and \( t_2 = -T. \)

\begin{align*}
\tau &= t_1 - t_2. \\
s &= t_1 + t_2. \\
t_1 &= \frac{1}{2}s + \frac{1}{2}\tau, \quad t_2 = \frac{1}{2}s - \frac{1}{2}\tau, \quad t_1 = T = \frac{1}{2}s + \frac{1}{2}\tau \rightarrow s = 2T - \tau. \quad t_1 = -T = \frac{1}{2}s + \frac{1}{2}\tau \rightarrow s = -2T - \tau. \\
\end{align*}

Figure 43.

\begin{align*}
J &= \left[ \begin{array}{cc} \frac{\partial t_1}{\partial \tau} & \frac{\partial t_2}{\partial \tau} \\ \frac{\partial t_1}{\partial \tau} & \frac{\partial t_2}{\partial \tau} \end{array} \right] = \left[ \begin{array}{cc} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{array} \right] = \frac{1}{2}. \\
E[|F_{X,T}(\omega)|^2] &= \frac{1}{2} \int_X \int R_X(\tau)e^{-j\omega\tau}d\tau ds. \\
E[|F_{X,T}(\omega)|^2] &= \frac{1}{2} \int_{-2T}^{0} R_X(\tau)e^{-j\omega\tau} \left[ \int_{-2T-\tau}^{-2T+\tau} d\sigma \right] d\tau + \frac{1}{2} \int_{0}^{2T} R_X(\tau)e^{-j\omega\tau} \left[ \int_{-2T-\tau}^{2T+\tau} d\tau \right] d\tau. \\
E[|F_{X,T}(\omega)|^2] &= 2T \left[ \int_{-2T}^{0} R_X(\tau)e^{-j\omega\tau} \left( 1 + \frac{\tau}{2T} \right) d\tau + \int_{0}^{2T} R_X(\tau)e^{-j\omega\tau} \left( 1 - \frac{\tau}{2T} \right) d\tau \right].
\end{align*}

Figure 44.
\begin{equation}
= 2T \int_{-2T}^{2T} R_X(\tau)e^{-j\omega \tau} \left(1 - \frac{|\tau|}{2T}\right) d\tau.
\end{equation}

\begin{equation}
\frac{E[|F_{X,T}(\omega)|^2]}{2T} = \int_{-2T}^{2T} F_X(\tau)e^{-j\omega \tau} \left(1 - \frac{|\tau|}{2T}\right) d\tau.
\end{equation}

\begin{equation}
\lim_{T \to \infty} \left(\frac{E[|F_{X,T}(\omega)|^2]}{2T}\right) = \lim_{T \to \infty} \int_{-2T}^{2T} R_X(\tau)e^{-j\omega \tau} \left(1 - \frac{|\tau|}{2T}\right) d\tau.
\end{equation}

\begin{equation}
\int_{-\infty}^{\infty} R_X(\tau)e^{-j\omega \tau} d\tau.
\end{equation}

\begin{equation}
\mathcal{F}\{R_x(\tau)\} = S_X(\omega).
\end{equation}


\begin{equation}
X(t) \quad \xrightarrow{\text{WSS}} \quad h(t) \quad \xrightarrow{\text{Y(t)}}
\end{equation}

\begin{equation}
E[Y(t)] = \int_{-\infty}^{\infty} h(t_1)E[x(t - t_1)]dt_1.
\end{equation}

\begin{equation}
= \mu_X \int_{-\infty}^{\infty} h(t_1)dt_1.
\end{equation}

\begin{equation}
= \mu_X H(0), \text{ constant}.
\end{equation}
\[
H(j\omega) = \int_{-\infty}^{\infty} h(t_1)e^{-j\omega t_1}dt_1.
\]

\[
H(0) = \int_{-\infty}^{\infty} h(t_1)dt_1.
\]

\[
E[Y(t + \tau)Y(t)] = E \left[ \int_{-\infty}^{\infty} h(v)x(t + \tau - v)dv \int_{-\infty}^{\infty} h(u)x(t - u)du \right].
\]

\[
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(v)h(u)R_X(\tau + u - v)dudv.
\]

\[
= \int_{-\infty}^{\infty} h(u) \left[ \int_{-\infty}^{\infty} h(v)R_X(\tau + u - v)dv \right] du.
\]

\[
= \int_{-\infty}^{\infty} h(u) \left( h * R_X \right)(\tau + u - v)du.
\]

\[
= \int_{-\infty}^{\infty} \tilde{h}(-u) \left( h * R_X \right)(\tau + u)du,
\]

where \( \tilde{h}(-u) = h(u) \). Let \( \theta = -u \) and \( d\theta = -du \)

\[
\Rightarrow E[Y(t + \tau)Y(t)] = -\int_{-\infty}^{\infty} \tilde{h}(\theta)(R_X * h)(\tau - \theta)d\theta.
\]

\[
= \int_{-\infty}^{\infty} \tilde{h}(\theta)(R_X * h)(\tau - \theta)d\theta.
\]

\[
= \tilde{h} * (R_X * h)(\tau) = R_Y(\tau).
\]

\[
\therefore Y(t) \text{ is WSS.}
\]

\[
S_Y(\omega) = \mathcal{F}\{R_Y(\tau)\} = H^*(\omega)H(\omega)S_X(\omega).
\]

\[
= |H(\omega)|^2S_X(\omega).
\]
23. Lecture 21: 11/12/03

23.1. **Linear Systems Response**. \( R \) is linear represented by an operator \( \theta : \mathcal{F}(t) \rightarrow \mathcal{F}(t) \).

\[ \theta_1 \text{ is linear:} \]

\[ (90) \]

\[ \delta(t-\tau) \xrightarrow{\theta_1} h_1(t,\tau) \]

**Figure 47.**

If \( \theta_1 \) is time-invariant then \( h_1(t, \tau) = h_1(t - \tau) \). \( \theta_2 \) is another linear system. Let

\[ (91) \]

\[ Y_1(t) = \theta_1(x)(t), \]

where \( x(t) \) has \( E[x(t)] = \mu_x(t) \).

\[ (92) \]

\[ Y_2(t) = \theta_2(x)(t). \]

\[ (93) \]

\[ R_{XX}(t_1, t_2) = E[x(t_1)x(t_2)]. \]

What is \( R_{Y_1Y_2}(t_1, t_2) \)?

\[ (94) \]

\[ R_{Y_1Y_2}(t_1, t_2) = E[Y_1(t_1)Y_2(t_2)]. \]

Corresponding to \( \theta_2 \), there is \( h_2(\cdot, \cdot) \).

\[ (95) \]

\[ R_{Y_1Y_2}(t_1, t_2) = E \left[ \int_{-\infty}^{\infty} h_1(t_1, \tau)x(\tau)d\tau \int_{-\infty}^{\infty} h_2(t_2, \eta)x(\eta)d\eta \right]. \]

If \( h_1(t, \tau) \) is the impulse response for \( \theta_1 \), then

\[ (96) \]

\[ \theta_1(x)(t) = \int_{-\infty}^{\infty} h(t, \tau)x(\tau)d\tau. \]

\[ (97) \]

\[ f(t) = \int_{-\infty}^{\infty} f(\tau)\delta(t - \tau)d\tau. \]

\[ (98) \]

\[ R_{Y_1Y_2}(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_1(t_1, \tau)h_2(t_2, \eta)R_{XX}(\tau, \eta)d\tau d\eta \]

\[ = \int_{-\infty}^{\infty} h_2(t_2, \eta) \left( \int_{-\infty}^{\infty} h_1(t_1, \tau)R_{XX}(\tau, \eta)d\tau \right) d\eta. \]

\[ (99) \]

\[ = \int_{-\infty}^{\infty} h_2(t_2, \eta) \theta_1(R_{XX}(\cdot, \eta))(t_1)d\eta. \]

\[ (100) \]

\[ = \theta_2(\theta_1(R_{XX}(\cdot, \cdot)))(t_1, t_2). \]
If $X$ is WSS, $\theta_1$ and $\theta_2$ are time invariant, then

$$R_{Y_1Y_2}(t_1, t_2) = (h_1(t_1) * R_{XX}(t_1, t_2)) * h_2(-t_2).$$

**Special Case:** If $\theta_1 = \theta_2$,

$$R_{YY}(t_1, t_2) = R_{YY}(t_1 - t_2).$$

$$= R_{YY}(\tau) = h(\tau) * h(-\tau) * R_{XX}(\tau).$$

$$S_{YY}(\omega) = |H(\omega)|^2 S_{XX}(\omega).$$

**More Special Cases:** If $\theta_1$ is the identity,

$$h_1(t - \tau) = \delta(t - \tau),$$

then

$$Y_1(t) = X(t)$$

and

$$Y_2(t) = \theta_2(x)(t).$$

$$R_{Y_1Y_2}(t_1, t_2) = h_2(-\tau) * R_{XX}(\tau).$$

If now, $\theta_2$ is the identity, then

$$Y_1(t) = \theta_1(x)(t)$$

and

$$Y_2(t) = X(t).$$

$$R_{YY}(\tau) = h(\tau) * h(-\tau) * R_{XX}(\tau) \quad (\theta_1 = \theta_2).$$
Example 34. (Problem 7.29 from textbook)

(114) \[ Y(t) = X(t) + 0.3 \frac{dX(t)}{dt}. \]

(115) \[ E[Y(t)] = E[X(t) + 0.3 \frac{dX(t)}{dt}]. \]

(116) \[ = \mu_X(t) + 0.3 E\left[ \frac{dX(t)}{dt} \right]. \]

(117) \[ = \mu_X(t) + 0.3 \frac{d}{dt} E[X(t)]. \]

(118) \[ \frac{dX(t)}{dt} = \lim_{n \to \infty} \left( x(t + \frac{1}{n}) - x(t) \right) / (1/n). \]

(119) \[ E\left[ \frac{dX(t)}{dt} \right] = E\left[ \lim_{n \to \infty} \left( \frac{x(t + \frac{1}{n}) - x(t)}{1/n} \right) \right]. \]

(120) \[ = \lim_{n \to \infty} E\left[ \frac{x(t + \frac{1}{n}) - x(t)}{1/n} \right]. \]

(121) \[ = \lim_{n \to \infty} \frac{\mu_X(t + \frac{1}{n}) - \mu_X(t)}{1/n}. \]

(122) \[ = \mu'_X(t). \]

To address switching the limit with “E”, we can generically look at the following problem:

\[ x_1, x_2, x_3, \ldots \text{ limit }_{n \to \infty} x_n = x \text{ in probability. Question: limit }_{n \to \infty} E[x_n] = E[x]? \]

Theorem 9. Bounded Convergence Theorem. If \( x_n \to x \) in probability and if \( |x_n| < B < \infty \), then

(123) \[ \lim_{n \to \infty} E[|x_n - x|] = 0. \]

Corollary 1. \( E[x_n] \to E[x] \).

Proof:

(124) \[ |E[x_n] - E[X]| = |E[x_n - x]| \leq E[|x_n - x|], \]

because \( E[Y] \leq E[|Y|] \).

(125) \[ \therefore \lim_{n \to \infty} |E[x_n - x]| \leq \lim_{n \to \infty} E[|x_n - x|] = \lim_{n \to \infty} E[|x_n - x|] = 0. \]
Proof of Theorem 9: \( x_n \) converges to \( X \) in probability. Pick \( \epsilon > 0 \), then there is an integer \( N \) such that

\[
P\{|x_n - x| > \epsilon\} < \epsilon, \forall n \geq N.
\]

\[
E[|x_n - x|] = E[|x_n - x|I_{\{|x_n - x| > \epsilon\}}] + E[|x_n - x|I_{\{|x_n - x| \leq \epsilon\}}].
\]

Note that \( |x_n - x| \leq |x_n| + |x| \leq B + B \) (assuming \( |x| < B \)).

\[
\therefore E[|x_n - x|I_{\{|x_n - x| > \epsilon\}}] \leq 2BE[I_{\{|x_n - x| > \epsilon\}}]
\]

\[
= 2BP\{|x_n - x| > \epsilon\} < 2B\epsilon.
\]

\[
E[|x_n - x|I_{\{|x_n - x| \leq \epsilon\}}] \leq \epsilon E[I_{\{|x_n - x| \leq \epsilon\}}]
\]

\[
= \epsilon P\{|x_n - x| \leq \epsilon\} \leq \epsilon.
\]

\[
\therefore E[|x_n - x|] \leq 2B\epsilon + \epsilon = (2B + 1)\epsilon.
\]

\[
\therefore \lim_{n \to \infty} E[|x_n - x|] = 0.
\]

Property 11. If \( f \) is a continuous function, then \( x_n \to x \) in probability implies \( f(x_n) \to f(x) \) in probability. In particular, if we take \( f(t) = t^2 \), then the bounded convergence theorem says that \( E[|x - x_n|^2] \to 0 \) as \( n \to \infty \) (\( \Rightarrow \) convergence in the mean-square sense).

Exercise 10. In problem 7.29, use the BCT to show that

\[
E[x'(t)] = \frac{d}{dt}E[x(t)].
\]

Example 35. On calculating \( R_X \): Let \( \theta \) be a uniformly distributed r.v. in \([0, \Delta]\).

\[
\tilde{x}(t) = \sum_{n=-\infty}^{\infty} (u(t - n\Delta) - u(t - (n + 1)\Delta))A_n,
\]

where \( A_n \in \{-1, 1\}, P\{A_n = 1\} = 1/2 \) and \( A_n \)'s are iid.

\( x(t) = \tilde{x}(t - \theta) \). \( E[x(t)] = 0 \). \( R_X(t, t + \tau) = E[x(t)x(t + \tau)] \). What happens when \( |\tau| > \Delta \)?

Then, \( R_{XX}(t, t + \tau) = 0, \forall t \). What if \( |\tau| \leq \Delta \)?

Then \( E[x(t)x(t + \tau)] = 1 \) if \( \theta \) is such that both \( x(t) \) and \( x(t + \tau) \) belong to the same bin. Look at the left edge for \( t \) (last “jump” before \( t \)), call it \( m\Delta + \theta \). Right edge: \( (m + 1)\Delta + \theta \). Requirement: \( m\Delta + \theta < t \). \( (m + 1)\Delta + \theta \geq t + \tau \). Namely: \( \theta < t - m\Delta \). \( \theta \geq t + \tau - (m + 1)\Delta \). \( t + \tau - (m + 1)\Delta \leq \theta < t - m\Delta \). \( \therefore \) the requirement occurs with probability

\[
\int_{t + \tau - (m + 1)\Delta}^{t - m\Delta} \frac{1}{\Delta} dx = \frac{1}{\Delta}(\Delta - \tau) = 1 - \frac{\tau}{\Delta}.
\]
If \( \tau \) is negative \( \Rightarrow (1 + \frac{\tau}{\Delta}) \).

\[
(137) \quad \therefore R_X(t, t + \tau) = \begin{cases} 
1 - \frac{|\tau|}{\Delta}, & |\tau| \leq \Delta \\
0, & |\tau| > \Delta 
\end{cases}
\]
24. Lecture 22: 11/17/03

24.1. Fact about Gaussian r.v.’s. If \( x_1, x_2, \ldots, x_n \) are individually Gaussian, then they are jointly Gaussian if and only if for every \( a_1, a_2, \ldots, a_n \),

\[
\sum_{i=1}^{n} a_i x_i \text{ is a Gaussian r.v.}
\]

Example 36. \( x_1 \) is Gaussian. \( x_2 = -x_1, x_1 + x_2 = 0 \), which is not a Gaussian r.v. \( \Rightarrow x_1 \) and \( x_2 \) are not jointly Gaussian.

24.2. Conditional Expectations. If \( x_1 \) and \( x_2 \) have joint pdf \( f_{x_1,x_2}(x_1,x_2) \), then

\[
E[x_1|x_2 = x_2] = \int x_1 f_{x_1|x_2}(x_1|x_2)dx_1,
\]

where

\[
f_{x_1|x_2}(x_1|x_2) = \frac{f_{x_1,x_2}(x_1,x_2)}{f_{x_2}(x_2)}
\]

and

\[
f_{x_2}(x_2) = \int_{-\infty}^{\infty} f_{x_1,x_2}(x_1,x_2)dx_1.
\]

24.3. Poisson Process. \( x_1, x_2, x_3, \ldots \) are iid, each exponentially distributed with mean \( \frac{1}{\lambda}, \lambda > 0 \).

\[
N(t) = \sum_{i=1}^{\infty} I_{[T_i \leq t]} = \sum_{i=1}^{\infty} n(t - T_i).
\]

Distribution of \( N(t) \): \( P\{N(t) = n\}, n = 0, 1, \ldots \)

\[
\{N(t) = n\} = \{T_n \leq t\} \cap \{T_{n+1} > t\}
\]

\[
= \{T_n \leq t, T_{n+1} > t\}
\]

\[
= \{T_n \leq t, x_{n+1} > t - T_n\}.
\]

To find \( P\{N(t) = n\} \), we first condition on the location of \( T_n \), call it \( \tau \).
We want \( x_{n+1} > t - \tau \) (\( \Rightarrow \) next arrival time falls beyond \( t \)). The range of \( \tau \) is \([0,t]\).

\[
P\{N(t) = n\} = \int_0^t f_{T_n}(\tau) d\tau P\{x_{n+1} > t - \tau\}.
\]

We can do this because \( x_{n+1} \) and \( T_n \) are independent.

\[
P\{x_{n+1} > t - \tau\} = \int_{t-\tau}^\infty f_{x_{n+1}}(x) dx = \int_{t-\tau}^\infty \lambda e^{-\lambda x} dx = e^{-\lambda(t-\tau)}.
\]

What is \( f_{T_n} \)?

\[
T_n = x_1 + x_2 + \ldots + x_n.
\]

\[
f_{T_n}(\tau) = f_{x_1}(\tau) * \ldots * f_{x_n}(\tau) = \frac{(\lambda\tau)^{n-1}\lambda}{(n-1)!}u(\tau).
\]

\[
P\{N(t) = n\} = \int_0^t \frac{\tau^{n-1}\lambda^n}{(n-1)!} e^{-\lambda\tau} e^{-\lambda(t-\tau)} d\tau
\]

\[
= \lambda^n e^{-\lambda t} \int_0^t \frac{\tau^{n-1}}{(n-1)!} d\tau
\]

\[
= \frac{(\lambda)^n e^{-\lambda t}}{n!}, \quad n = 0, 1, 2, \ldots.
\]

\( N(t) \) is a Poisson r.v. with mean \( \lambda t \), where \( \lambda \) is the average rate of arrival. \( N(t) \) is called a Poisson process. \( E[N(t)] = \lambda t \). We think of \( \lambda \) as the average rate of arrival. \( E[N^2(t)] = (\lambda t) + (\lambda t)^2 \) because \( \text{var}(N(t)) = \lambda t \).

Signal to Noise Ratio (SNR):

\[
\text{SNR} = \frac{E[N(t)]}{\sigma_N(t)} = \frac{\lambda t}{\sqrt{\lambda t}} = \sqrt{\lambda}.
\]

\( \sigma_N(t) \) is the uncertainty. As \( t \uparrow \), the uncertainty \( \searrow \) (variability).

\[
R_{NN}(t_1,t_2) = E[N(t_1)N(t_2)]; \quad t_1 \leq t_2
\]

\[
= E[N(t_1)\{N(t_2) - N(t_1) + N(t_1)\}]
\]

\[
= E[N(t_1)\{N(t_2 - N(t_1)) + N(t_1)\}]
\]
\[ (157) \quad t^2 t_1 t_2 = \lambda t_1 + \lambda^2 t_1 t_2. \]

\[ (158) \quad K_{NN}(t_1, t_2) = R_{NN}(t_1, t_2) - E[N(t_1)|E[N(t_2)]] \]

\[ (160) \quad K_{NN}(t_1, t_2) = \lambda t_1 + \lambda^2 t_1 t_2 = \lambda. \]

\[ (162) \quad K_{NN}(t_1, t_2) = \lambda \min(t_1, t_2). \]

24.4. **Hazard or Failure Rates.** Consider \( x_1 = T_1 \). Mean arrival time of \( 1 \). If we know that \( x_1 > x \), what is the probability that it comes in \( [x, x + \Delta x] \)? The probability is:

\[ (164) \quad P\{x < x_1 \leq x + \Delta x|x_1 > x\} = \frac{P\{x < x_1 \leq x + \Delta x, x_1 > x\}}{P\{x_1 > x\}} \]

\[ (165) \quad \frac{P\{x < x_1 \leq x + \Delta x, x_1 > x\}}{1 - P\{x_1 \leq x\}} \approx \frac{f_{x_1}(x)\Delta x}{1 - F_{x_1}(x)} \]

\[ (166) \quad \lim_{\Delta x \to 0} \frac{P\{x < x_1 \leq x + \Delta x|x_1 > x\}}{\Delta x} = \frac{f_{x_1}(x)}{1 - F_{x_1}(x)} \equiv a(x). \]

If \( f_{x_1}(x) = \lambda e^{-\lambda x}u(x) \) then

\[ (168) \quad a(x) = \frac{\lambda e^{-\lambda x}u(x)}{e^{-\lambda x}u(x)} = \lambda. \]

Conclusion: \( \lambda \), which we interpreted as a mean arrival rate, is actually a failure rate (hazard rate). That is, \( \lambda = a \left( a = \frac{f_{x_1}(x)}{1 - F_{x_1}(x)} \right). \)

Conversely: If we know the failure rate of some phenomenon, call it \( a(x) \), what is the density function \( f_{x_1}(x) \)? \( a(x) = \frac{f_{x_1}(x)}{1 - F_{x_1}(x)}. \)

\[ (169) \quad \int_0^x a(u)du = \int_0^x \frac{f(u)}{1 - F(u)}du = -\log(1 - F(u))|_0^x \]

\[ = -\log(1 - F(x)) + \log(1 - F(0)) = -\log(1 - F(x)). \]

\[ (171) \quad \Rightarrow 1 - F(x) = e^{-\int_0^x a(u)du}. \]
\[ F(x) = 1 - e^{-\int_0^x a(u)du} \]

and

\[ f(x) = a(x)e^{-\int_0^x a(u)du}. \]

**Example 37.** If \( a(x) = a \), then \( f_a(x) = ae^{-ax}u(x) \).

\[ a(x) = \begin{cases} 
\lambda, & x \geq x_d \\
0, & x < x_d
\end{cases} \]

\[ \Rightarrow f_x(x) = \lambda e^{-\lambda(x-x_d)}u(x-x_d). \]

**Figure 54.**

**Figure 55.** Some fail with probability \( p \). \( \mu(n) \) is the average time % load \( n \).
25.1. Linear Estimation

\[ s \xrightarrow{\text{Distortion + Noise}} x_i \]

signal (r.v.) indexed by \( i = 1, 2, \ldots, n \)

**Question:** Estimate \( s \) from \( x_1, x_2, \ldots, x_n \). A simple approach: Form a linear function of the data (linear estimator). What is the best linear estimator that we can get? In what sense? In the sense of minimizing the mean-square error (MSE).

**Problem:** Find the coefficients, \( a_1, \ldots, a_n \), such that

\[ \text{MSE} = E \left[ \left( s - \sum_{i=1}^{n} a_i x_i \right)^2 \right] \]

is minimized.

**Solution:** Method: Take \( \frac{d \text{MSE}}{da_j} = 0, \forall j \).

\[ \frac{d}{da_j} \left( E \left[ \left( s - \sum_{i=1}^{n} a_i x_i \right)^2 \right] \right) = E \left[ \frac{d}{da_j} \left( s - \sum_{i=1}^{n} a_i x_i \right)^2 \right] 
\]

\[ = E \left[ -2 \left( s - \sum_{i=1}^{n} a_i x_i \right) x_j \right] = 0. \]

\[ \therefore E[sx_j] = E \left[ \sum_{i=1}^{n} a_i x_i x_j \right]. \]

Call \( E[sx_j] = R_{o,j} \). \( R_{ij} = E[x_i x_j] \), for any \( j \). We can write it in matrix form: \( R_o = \tilde{a}R \), where \( \tilde{R}_o = [R_{o1} R_{o2} \ldots R_{on}] \), for any \( j \). \( \tilde{a} = [a_1 \ldots a_n] \). We had seen that unless the \( x_i \)'s are degenerate, \( R \) is positive definite (hence invertible). \( \therefore \tilde{a} = \tilde{R}_o R^{-1} \).

**Example 38.** Discrete Case: \( x_i = S + N_i, i = 1, 2, \ldots, n \). \( \hat{s} = a_1 x_1 + \ldots + a_n x_n \). \( R_o = aR \).

\[
R = \begin{bmatrix}
\sigma_N^2 + s^2 & s^2 & \ldots \\
 s^2 & \sigma_N^2 + s^2 & \ldots \\
 \ldots & \ldots & \ddots 
\end{bmatrix}.
\]

\[ R_o = \begin{bmatrix}
 s^2 \\
 \vdots \\
 s^2 
\end{bmatrix}.\]
Recall that we had:

\[ E[sX_j] = E \left[ \sum_{i=1}^{n} a_i x_i x_j \right]. \]  

(188)  

\[ E \left[ x_j \left( s - \sum_{i=n} a_i x_i \right) \right] = 0, \quad \forall j. \]  

(189)  

Error is orthogonal Principle.

25.2. Linear Estimation: Continuous Case.

\[ \hat{s}(t) = \int_a^b X(\alpha) h_t(\alpha) d\alpha. \]  

(190)  

This is a linear estimate of \( s(t) \) using the data in \([a, b] \).

Question: What is the best \( h_t(\alpha) \) that minimizes \( MSE = E[(s(t) - \hat{s}(t))^2] \) for a fixed \( a, b, t \).

Answer: Such an \( h_t(\alpha) \) must satisfy the orthogonality principle. That is

\[ E \left[ (s(t) - \int_a^b x(\alpha) h_t(\alpha) d\alpha) x(\tau) \right] = 0, \quad \forall \tau \in [a, b]. \]  

(191)
\( \bar{R}_o = \bar{a}R \) is called the Wiener-Hopf Equation.

Suppose that we are interested in a time-invariant filter \( h(t) \). That is, \( h(t) = h(t - \cdot) \). Then the Wiener-Hopf equation becomes

\[
E \left[ \left( s(t) - \int_a^b h(t - \alpha)x(\alpha)d\alpha \right)x(\tau) \right] = 0, \quad \tau \in [a, b].
\]

\[
\therefore R_{SX}(t, \tau) = \int_a^b h(t - \alpha)R_{XX}(\alpha, \tau)d\alpha, \quad \tau \in [a, b].
\]

If \( s(t) \) is WSS, \( x(t) \) is WSS, then \( \hat{s} \) is also WSS.

\[
R_{SX}(t, \tau) = R_{SX}(t - \tau).
\]

\[
R_{XX}(\alpha, \tau) = R_{XX}(\alpha - \tau) = R_{XX}(\tau - \alpha).
\]

\( \therefore \) with stationarity assumptions:

\[
R_{SX}(t - \tau) = \int_a^b h(t - \alpha)R_{XX}(\tau - \alpha)d\alpha, \quad \forall \tau \in [a, b] \text{ (Wiener-Hopf equation)}.
\]

Special case: Non-causal Wiener filter. \( a = -\infty, b = +\infty \).

\[
R_{SX}(t, \tau) = \int_{-\infty}^\infty h(t - \alpha)R_{XX}(\tau - \alpha)d\alpha, \quad \tau \in (-\infty, \infty).
\]

Pick \( \tau = 0 \) (no loss of anything).

\[
R_{SX}(t) = \int_{-\infty}^\infty h(t - \alpha)R_{XX}(\alpha)d\alpha = h(t) * R_{XX}(t).
\]

\[
S_{SX}(\omega) = H(\omega)S_{XX}(\omega).
\]

\[
H(\omega) = \frac{S_{SX}(\omega)}{S_{XX}(\omega)} \quad \text{Non-causal Wiener filter.}
\]

Example 39.

\[
H(\omega) = \frac{S_{SX}(\omega)}{S_{XX}(\omega)} = ?
\]

**Observation Model**

\[ s(t) \rightarrow \bullet \rightarrow x(t) \rightarrow \hat{s}(t) \]

WSS

Data

\[ v(t) \]

White Gaussian Noise (WGN)

**Reconstruction (Restoration)**

\[ x(t) \rightarrow [h(t)] \rightarrow \hat{s}(t) \]

Figure 58.
Find $R_{SX}(\tau)$ first:

\begin{align}
R_{SX}(\tau) &= E[s(t)x(t + \tau)] \\
&= E[s(t)(s(t + \tau) + v(t))]
\end{align}

which is assumed to be known.

\begin{align}
S_{XX}(\omega) &= \mathcal{F}\{R_{SS}(\tau) + R_{VV}(\tau)\} = S_{SS}(\omega) + 1 \\
\Rightarrow H(\omega) &= \frac{S_{SS}(\omega)}{1 + S_{SS}(\omega)}.
\end{align}

**Example 40.**

\begin{align}
H(\omega) &= \frac{S_{SX}(\omega)}{S_{VV}(\omega)}.
\end{align}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure59.png}
\caption{Figure 59.}
\end{figure}

\begin{align}
R_{SX}(\tau) &= E[s(t)x(t + \tau)] \\
&= E[s(t)((s * g)(t + \tau) + v(t + \tau))] \\
&= E\left[ s(t) \int_{-\infty}^{\infty} s(t + \tau - \alpha)g(\alpha)d\alpha \right] \\
&= R_{SS}(\tau) * g(-\tau). \\
\therefore S_{SX}(\omega) &= G(\omega)S_{SS}(\omega).
\end{align}

Similarly,

\begin{align}
S_{XX}(\omega) &= |G(\omega)|^2S_{SS}(\omega) + S_{VV}(\omega).
\end{align}
\[(214) \quad H(\omega) = \frac{G * (\omega) S_{SS}(\omega)}{|G(\omega)|^2 S_{SS}(\omega) + S_{VV}(\omega)}.
\]

\[(215) \quad R_{XX}(\tau) = R_{SS}(\tau) * g(-\tau) * g(\tau).
\]
26. Lecture 24: 11/24/03

26.1. **Linear Filtering.** Non-causal optimal filter:

\[
\hat{s}(t) = \int_{-\infty}^{\infty} h_{nc}(t-\tau)x(\tau)d\tau.
\]

x: data, \(R_{XX}\) known. s: signal, \(R_{SS}\) known, \(R_{SX}\) known. If

\[
H_{nc}(\omega) = \frac{S_{SX}(\omega)}{S_{XX}(\omega)},
\]

then

\[
E^2 = E[(\hat{s}(t) - s(t))^2]
\]

is minimized over all other LTI filters \(h(\cdot)\). We used an orthogonality principle: Error \(\perp\) data used to generate estimate, i.e.,

\[
E[(s(t) - \hat{s}(t))] = 0, \forall \tau.
\]

26.2. **Causal Wiener Filtering.** We want:

\[
\hat{s}(t) = \int_{-\infty}^{\infty} h(t-\tau)x(\tau)d\tau.
\]

So, we want \(h(\cdot)\) that minimizes the MSE,

\[
E^2 = E[(\hat{s}(t) - s(t))^2].
\]

Call the filter \(h(\cdot)\). Use the orthogonality principle:

\[
E[(s(t) - \hat{s}(t))x(\tau)] = 0, \forall \tau \leq t.
\]

\[
\therefore R_{SX}(t-\tau) = E[\hat{s}(t)x(\tau)]
\]

\[
= E \left[ \int_{-\infty}^{\infty} h(t-\alpha)x(\alpha)d\alpha x(\tau) \right]
\]

\[
= \int_{-\infty}^{t} h(t-\alpha)R_{XX}(t-\alpha)x(\tau) d\alpha, \tau \leq t.
\]

Change of variable: \(u = t - \alpha\).

\[
\Rightarrow R_{SX}(t-\tau) = \int_{0}^{\infty} h(u)R_{XX}(t-u-\tau)du, \tau \leq t.
\]

True when \(\tau = 0\). Call \(\eta = t - \tau\).

\[
R_{SX}(\eta) = \int_{0}^{\infty} h(u)R_{XX}(\eta-u)du, \eta \geq 0
\]

\[
= \int_{-\infty}^{\infty} h(u)R_{XX}(\eta-u)du, \eta \geq 0.
\]
(Looks like convolution, but only for positive $\eta$, it’s not a convolution $\Rightarrow$ cannot apply Fourier Transform). For the moment, suppose $R_{XX}(\tau) = \delta(\tau)$. Equation (228) gives: $R_{SX}(\tau) = h(\tau)$, $\tau \geq 0$, but $h(\tau) = 0$, if $\tau < 0$ by causality.

Real

\[ h_c(t) = \gamma(t) \ast (R_{SX}(t)u(t)). \]

\[ h_c(t) = \gamma(t) \ast [R_{SY}(t) + \gamma(-t))u(t)]. \]

Error Calculation:

\[ \text{MSE} = E[(\hat{s}(t) - s(t))^2] = E[(\hat{s}(t) - s(t))(\hat{s}(t) - s(t))] \]

\[ = -E[(\hat{s}(t) - s(t)s(t)] \]

because $\hat{s}(t)$ is orthogonal to $(\hat{s}(t) - s(t))$.

\[ \text{MSE} = R_{SS}(0) - E[s(t)\hat{s}(t)] \]

\[ = R_{SS}(0) - E[s(t)\hat{s}(t)] \]

\[ = R_{SS}(0) - E\left[s(t)\int_{-\infty}^{t} x(u)R_{SX}(t-u)du\right] \]

\[ = R_{SS}(0) - \int_{-\infty}^{t} R_{SX}(t-u)R_{SX}(t)du. \]

$y = t - u.$

\[ = R_{SS}(0) - \int_{0}^{\infty} R_{SX}^2(y)dy \]
What about $\gamma(t)$? $S_z$

$Y(t) \rightarrow \gamma(t) \rightarrow x(t)$

**white**

Figure 61.

$$\frac{1}{S_{YY}(\omega)} = \Gamma(j\omega)\Gamma(-j\omega).$$

**Example 41.**

$$Y(t) = s(t) + v(t),$$

where $Y(t)$ is data, $s(t)$ is WSS, $v(t)$ is WGN and $s$ and $v$ are uncorrelated.

$$S_{SS}(\omega) = \frac{12}{4 + \omega^2}.$$

$S_{VV}(\omega) = 1 \ (w/\text{rad/s}).$

Find $h_c(t)$ and the error.

$$h_c(t) = \gamma(t) * [R_{SY}(t) * \gamma(-1)]u(t).$$

$$S_{YY}(\omega) = S_{SS}(\omega) + S_{VV}(\omega)$$

$$= \frac{16 + \omega^2}{4 + \omega^2} = \frac{(4 + j\omega)(4 - j\omega)}{(2 + j\omega)(2 - j\omega)}$$

$$= \Gamma^{-1}(j\omega)\Gamma^{-1}(-j\omega).$$

$$\therefore \Gamma(j\omega) = \frac{2 + j\omega}{4 + j\omega} : \text{causal.}$$

$$\Gamma(s) = \frac{2 + j\omega}{4 + j\omega} = 1 - \frac{2}{s + 4}.$$  

$$\gamma(t) = s(t) - 2e^{-4t}u(t) : \text{causal.}$$

**Compute $R_{SY}(t)$**.

$$R_{SY}(t) = R_{SS}(t) = \mathcal{F}^{-1}\left\{ \frac{12}{4 + \omega^2} \right\}$$

$$\mathcal{L}^{-1}\left\{ \frac{12}{4 - s^2} \right\} = \mathcal{L}^{-1}\left\{ \frac{3}{s + 2} + \frac{3}{s - 2} \right\}. $$
\[ R_{SY}(t) \ast \gamma(-t), \text{i.e., } S_{SY}(s)\Gamma(-s), \]

(252) \[ = \mathcal{L}^{-1}\{S_{SY}(s)\Gamma(-s)\} \]

(253) \[ = \mathcal{L}^{-1}\left\{ \frac{12}{4-s^2} \frac{2-s}{4-s} \right\} \]

(254) \[ = \mathcal{L}^{-1}\left\{ \frac{12}{(2-s)(s+2)} \frac{2-s}{4-s} \right\} \]

(255) \[ (R_{SY}(t) \ast \gamma(-t))u(t) = \mathcal{L}^{-1}\left\{ \frac{12}{(2+s)(4-s)} \right\} \text{ : causal} \]

(256) \[ = \mathcal{L}^{-1}\left\{ \frac{2}{2+s} + \frac{2}{4-s} \right\} \]

(257) \[ = 2e^{-2t}u(t). \]

(258) \[ h_c(t) = [(R_{SY}(t) \ast \gamma(-t))u(t)] \ast \gamma(t) = 2e^{-2t}u(t) \ast (\delta(t) - 2e^{-4t}u(t)). \]

(259) \[ H_c(s) = \frac{2}{s+2} \left[ \frac{2+s}{4+s} \right] = \frac{2}{4+s}. \]

(260) \[ \therefore h_c(t) = 2e^{-4t}u(t). \]

(261) \[ MSE = R_{SS}(0) - \int_0^\infty [R_{SX}(t)]^2 dt \]

(262) \[ = 3 - \int_{-\infty}^\infty (2e^{-2t}u(t))^2 dt = 2 \]

(263) \[ R_{SS}(\tau)F^{-1}\left\{ \frac{12}{4+\omega^2} \right\} = 3e^{-2|\tau|}. \]
27.1. Spectral Factorization

the output of an LTI system wi

\[ W(t) \xrightarrow{\frac{N_0}{2} \text{ (w/rads)}} H \xrightarrow{} Y(t) \]

**Figure 62.** $H$ is a ratio of polynomials.

\[
S_{YY}(\omega) = |H(\omega)|^2 \frac{N_0}{2}.
\]

We can write:

\[
H(\omega) = \frac{A(\omega^2) + j\omega B(\omega^2)}{C(\omega^2) + jD(\omega^2)},
\]

where $A, B, C$ and $D$ are polynomials with real coefficients.

\[
\therefore |H(\omega)|^2 = \frac{A^2(\omega^2) + \omega^2 B^2(\omega^2)}{C^2(\omega^2) + \omega^2 D^2(\omega^2)}.
\]

\[
\therefore |H(\omega)|^2 \text{ is a ratio of two polynomials that are even and have real coefficients.}
\]

**Consequence 1:** Because the coefficients are real, zeros and poles of $|H|^2$ occur with conjugate symmetry if complex. That is, if $s = \sigma + j\omega$ is a pole (zero), then so is $s^* = \sigma - j\omega$.

**Consequence 2:** Be

\[
s(\omega) = S(s)|_{s=j\omega} = \prod_p (s - sp) \prod_k (s - s_k)
\]

\[
\therefore \text{ we can write:}
\]

\[
(267) \quad s(\omega) = S(s)|_{s=j\omega} = \prod_p (s - sp) \prod_k (s - s_k)
\]
where $\phi^+(s) = \phi^-(s)$ and $S(\omega) = |\phi^+(\omega)|^2 = |\phi^-(\omega)|^2$. In particular, $\phi^+$ is stable and $\frac{1}{\phi^+}$ is also stable.

**Example 42.**

$$S(\omega) = \frac{2\omega^2 + 3}{\omega^2 + 1} = \frac{\sqrt{2}(j\omega + \sqrt{2/3})(j\omega - \sqrt{2/3})\sqrt{2}}{(j\omega + 1)(j\omega - 1)}.$$ 

(269) 

$$\phi^+(s) = \frac{\sqrt{2}(s + \sqrt{2/3})}{s + 1}.$$ 

(270) 

$$\phi^-(s) = \frac{\sqrt{2}(s - \sqrt{2/3})}{s - 1}.$$ 

27.2. **Optimal Linear Prediction.** We want the following: $\lambda > 0$. Estimate $s(t + \lambda)$ using the data $x(\tau)$ in the range $\tau \in [-\infty, t]$. We do it in a linear fashion:

(272) 

$$\hat{s}(t + \lambda) = \int_0^\infty x(t - \alpha)h(\alpha)d\alpha,$$

where $h(\alpha)$ is a causal filter. Equivalently,

(273) 

$$\hat{s}(t) = \int_0^\infty x(t - \lambda - \alpha)h(\alpha)d\alpha.$$ 

We can think of this equation as a causal Wiener filter but with delayed data (delay is a linear transformation). What is the optimal $h$? We appeal to the orthogonality principle: error $\perp$ data.

(274) 

$$e(t + \lambda) = \hat{s}(t + \lambda) - s(t + \lambda).$$ 

(275) 

$$E[(\hat{s}(t + \lambda) - s(t + \lambda))x(\alpha)] = 0, \; \forall \alpha \leq t.$$ 

We obtain:

(276) 

$$R_{SX}(t + \lambda - \alpha) = R_{\hat{S}X}(t + \lambda - \alpha), \; \forall \alpha \leq t.$$ 

(277) 

$$R_{\hat{S}X}(\tau) = E\left[\int_0^\tau h(\alpha)x(\tau - \lambda - \alpha)d\alpha(x(\tau)\right].$$ 

(278) 

$$R_{\hat{S}X}(\tau + \lambda) = \int_0^\infty h(\alpha)R_{XX}(\tau - \alpha)d\alpha.$$ 

Let $\beta = t - \alpha$. This gives:

(279) 

$$R_{SX}(\lambda + \beta) = R_{\hat{S}X}(\lambda + \beta), \; \beta \in [0, \infty)$$ 

(280) 

$$= \int_0^\infty h(\alpha)R_{XX}(\beta - \alpha)d\alpha.$$
\[
R_{SX}(\beta) = \int_0^\infty h(\alpha)R_{XX}(\beta - \lambda - \alpha) d\alpha.
\]

Solution:

\[
h_w(t) = \gamma(t) * [R_{SY}(t + \lambda) * \gamma(-t)]u(t).
\]

The \(+\tau\) term is the difference with causal \(h(t)\).

![Diagram](image-url)

**Form:** \(R_{SY}(t + \lambda) * \gamma(-t)\)

**Figure 64.**

27.3. **Smoothing.** \(\lambda < 0\).

\[
s(t + \lambda) = \int_0^\infty h(\alpha)x(t - \alpha) d\alpha.
\]

In this case, we move \(R_{SY}\) to the right by \(\lambda\). Let \(x'\) be a delayed version of \(x\). The ingredient in the causal Wiener filter was \(R_{SX}\) (we can use \(R_{SX}\)), we only care about correlation between \(s\) and data. But, \(R_{SX}(\tau) = R_{SX}(\tau) * w(-\tau)\), where \(w(\tau) = \delta(t - \lambda) : \text{delay. Linear estimation: Filtering, prediction and smoothing.}\)

27.4. **Markov Chains.** \(x_0 \ x_1 \ldots \ x_n \ldots\) (discrete valued) is a Markov chain if

\[
P\{x_n = j | x_1 = i_1, x_2 = i_2, \ldots, x_{n-1} = i_{n-1}\} = P\{x_n = j | x_{n-1} = i_{n-1}\}.
\]

**Example 43.** Gambling: Initial fortune = \(x_0 = X_0\). Each time we either increment our fortune or decrement (with probability \(p\)). Let \(x_n\) denote the fortune at time \(n\).

\[
P\{x_n = j | x_1 \ldots x_{n-1}\} = P\{x_n = j | x_{n-1}\}.
\]
Problem: Suppose that we start gambling if we reach a goal. \( X_o = x_o, \ x_{\text{max}} = \text{goal.} \) You also stop whenever you lose all of your money. Let \( P(x) \) be the probability of achieving your goal, where \( x \) is the initial fortune. Also, \( p \) is the probability of winning a hand. Recipe for \( P(x) \): \( P(0) = 0. \)

\[
P(x) = pP(x + 1) + (1 - p)P(x - 1). \quad (286)
\]

\[
x = \{0, 1, 2, \ldots , x_{\text{max}}\}. \quad (287)
\]
28. Lecture 26: 12/01/03

28.1. Linear Prediction. \( X(t) \) WSS data, conveys information on \( s(t) \) (WSS). We want to predict \( s(t + \lambda) \) using data in the range \((-\infty, t)\).

\[
\hat{s}(t + \lambda) = \int_0^\infty h(\alpha)x(t - \alpha)d\alpha.
\]

Orthogonality principle:

\[
e(t + \lambda) \perp x(t - \alpha), \quad \alpha \geq 0.
\]

\[
e(\cdot) = s(\cdot) - \hat{s}(\cdot).
\]

\[
E[\{\hat{s}(t + \lambda) - s(t + \lambda)\}x(t - \alpha)] = 0, \quad \forall \alpha \geq 0.
\]

\[
\int_0^\infty h(\beta)R_{XX}(\alpha - \beta)d\beta = R_{SX}(\lambda + \alpha), \quad \alpha \geq 0.
\]

If \( x \) was white \( (R_{XX}(\alpha - \beta) = \delta(\alpha - \beta)) \), then

\[
\hat{s}(t + \lambda) = \gamma(t) \star u(t) \star \gamma(t).
\]

\[
\begin{array}{c}
x \rightarrow \gamma(t) \rightarrow I \rightarrow h_i \rightarrow \hat{s}(t + \lambda) \\
\text{White} \quad \text{h(t)}
\end{array}
\]

Figure 65.

But,

\[
R_{ST}(t) = R_{SX}(t) \ast \gamma(-t).
\]

\( \therefore \) the total

\[
h(t) = h_1(t) \ast \gamma(t).
\]

Recipe:

\[
h(t) = \{(R_{SX} \ast \gamma(-\cdot))(\lambda + t)u(t)\} \ast \gamma(t).
\]

Error:

\[
e = E[e^2(t + \lambda)] = -E[\{\hat{s}(t + \lambda) - s(t + \lambda)\}s(t + \lambda)]
\]
\[ (298) \quad ECE 541 \quad 91 \]
\[ = R_S(0) - E \left[ \int_0^\infty h(\alpha)x(t-\alpha)s(t+\lambda)d\alpha \right] \]
\[ (299) \quad = R_S(0) - \int_0^\infty h(\alpha)R_{SX}(\lambda + \alpha)d\alpha \]
\[ (300) \quad = R_{SS}(0) - \int_0^\infty h_0^2(\alpha)d\alpha. \]

**Example 44.**

\[ (301) \quad S_{SS}(s) = \frac{49 - 25s^2}{(1-s^2)(9-s^2)} = S_{XX}(s). \]

No noise:

\[ (302) \quad S_{SS}(s) = \frac{(7 + 5s)(7 - 5s)}{(1+s)(3+s)(1-s)(3-s)}. \]

\[ (303) \quad \Gamma(s) = \frac{(1+s)(3+s)}{(7 + 5s)}. \]

\[ (304) \quad h_I(t) = R_{SI}(\lambda + t)u(t) = [R_{SX}(\lambda + t) \ast \gamma(\lambda + t)]u(t) \]
\[ (305) \quad = [R_{SS}(\lambda + t) \ast \gamma(-(\lambda + t))] \ast u(t). \]

In our case, \( R_{SX} = R_{SS}. \)

\[ (306) \quad S_{SI}(s) = S_{SS}(s)\Gamma(-s) = \frac{1}{\Gamma(s)}. \]

Recall that

\[ (307) \quad h_I(t) = R_{SI}(\lambda + t)u(t). \]

Put

\[ (308) \quad \mathcal{L}^{-1} \left\{ \frac{1}{\Gamma(s)} \right\} = \sum_i c_i e^{s_i t}u(t). \]

\[ (309) \quad R_{SI}(\lambda + t) = \sum_i c_i e^{s_i(t+\lambda)}u(t). \]

\[ (310) \quad u(t)R_{SI}(\lambda + t) = \sum_i c_i e^{s_i(t+\lambda)}u(t). \]

\[ (311) \quad H_I(s) = \frac{c_1}{s - s_1} e^{s_1 \lambda} + \frac{c_2}{s - s_2} e^{s_2 \lambda} \bigg|_{s_1 = 1}^{s_1 = -1} \bigg|_{c_1 = 1}^{c_2 = 4}. \]

\[ (312) \quad H_I(s) = \frac{e^{-\lambda}}{s + 1} + \frac{4e^{-3\lambda}}{s + 3} \bigg|_{\lambda = \log 2}. \]

\[ (313) \quad = \frac{s + 2}{(s + 1)(s + 3)}. \]
(314) \[ h(t) = \gamma(t) * h_I(t). \]

(315) \[ H(s) = \Gamma(s)H_I(s). \]

(316) \[ \frac{(1 + s)(3 + s)(s + 2)}{(7 + 5s)(s + 1)(s + 3)} = \frac{s + 2}{7 + 5s}. \]

(317) \[ h(t) = 0.2\delta(t) + \frac{3}{25}e^{-1.4t}u(t). \]

(With noise, we will have a different \( R_{SI} \).)

28.2. Markov Chains. Let

\[ Q(x) = pQ(x + 1) + qQ(x + 1), \]

where \( p \) is the probability of winning the first hand and \( q \) is the probability of losing the first hand. \( Q(0) = 0. \ Q(x_{max}) = 1. \) Propose a solution of the form \( r^x \) and plug it in.

(318) \[ r^x = pr^{x+1} + qr^{x-1}. \]

(319) \[ r^{x+1} = pr^{x+2} + qr^x. \]

(320) \[ r^2 + \frac{q}{p} - \frac{r}{p} = 0. \]

(321) \[ r^2 + \frac{q}{p} - r = 0. \]

(322) \[ r_1,2 = \frac{1 \pm \sqrt{1 - 4p(1 - p)}}{2p}. \]

If \( r_1 \neq r_2, \)

(323) \[ Q(x) = Ar_1^x + Br_2^x. \]
Apply the initial conditions,

\[ Q(x) = \frac{1}{r_1 x_{\text{max}} - r_2 x_{\text{max}}} (r_1 x - r_2 x_{\text{max}}) \].

1 = r_1 = r_2 \text{ when } p = \frac{1}{2}.

\[ Q(x) = A + Bx. \]

Apply initial conditions:

\[ Q(x) = \frac{x}{x_{\text{max}}}. \]

Let \( x_n \) denote the fortune at time \( n \).

\[ \{\text{winning}\} = \{x_i = x_{\text{max}}, \forall i\}. \]

Also, if winning occurs, then

\[ \{x_i = x_{\text{max}}, \text{infinitely many times occurs}\}. \]

\[ \{\text{winning}\} = \{x_i = x_{\text{max}}, \text{infinitely often}\}. \]

\[ \{\text{loose}\} = \{x_i = 0, \text{infinitely often}\}. \]

We just proved that

\[ P\{x_i = x_{\text{max}}, \text{infinitely often}\} = \frac{x}{x_{\text{max}}}. \]

\[ P\{x_i = 0, \text{infinitely often}\} = 1 - \frac{x}{x_{\text{max}}}, \]

by doing the same analysis, for the probability of losing as we did for \( Q(x) \). These two imply that

\[ P\left\{ \lim_{n \to \infty} x_n = x_{\text{max}} \right\} = \frac{x}{x_{\text{max}}}. \]

\[ P\left\{ \lim_{n \to \infty} x_n = 0 \right\} = 1 - \frac{x}{x_{\text{max}}}. \]

\[ P\left\{ \lim_{n \to \infty} x_n \neq 0 \right\} = \frac{x}{x_{\text{max}}}. \]

\[ \Rightarrow x_n \to x \text{ almost surely}. \]
28.3. **Martingales.** A sequence \( x_1, x_2, \ldots \) is a martingale if \( E[x_{n+k}|x_n] = x_n \). If we consider the gambler’s example, the fortune “\( x_n \)” is a martingale if \( p = \frac{1}{2} \). If the condition \( E[x_{n+k}|x_n] = x_n \) is replaced by \( \leq (\geq) \), we call \( x_n \) a super (sub) martingale.

**Theorem 10. Martingale Convergence Theorem.** If \( x_n \) is a sub-, super-, or martingale and \( E[|x_1|] < \infty \), then \( x_n \) converges to a r.v. almost surely.

In real numbers, if \( \{x_n\} \) with a property that \( |x_n| < B < \infty, x_{n+1} \geq x_n \), then \( x_n \) is convergent (from Completeness of Real Numbers, the Bolzano-Weirstrass Theorem).

\[
E[x_{n+1}|x_n] \geq x_n \Rightarrow E[x_{n+1}] \geq E[x_n].
\]

28.4. **Markov Chains.** Back to the gambling problem.

\[
P_{ij} = P\{x_{n+1} = j|x_n = i\} = \begin{cases} p, & \text{if } j = i + 1, \ i \neq x_{\max} \\ q, & \text{if } j = i - 1, \ i \neq 0 \\ 1, & \text{if } i = j = x_{\max} \\ 1, & \text{if } i = j = 0 \end{cases}
\]

Define the probability transition matrix \( P = ((P_{ij})) \). For our example:

\[
P = \begin{bmatrix} 1 & 0 & 0 & \ldots & \ldots & \ldots & \ldots & \ldots & 0 \\ q & 0 & p & 0 & \ldots & \ldots & \ldots & \ldots & : \\ 0 & q & 0 & p & 0 & \ldots & \ldots & \ldots & : \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \ldots & \ldots & q & 0 & p \\ 0 & 0 & \ldots & \ldots & \ldots & \ldots & q & 0 \end{bmatrix}.
\]

\( P \) has the property that the sum of any row is 1, which is called a stochastic matrix.

**Exercise 11.** Show that \( \lambda = 1 \) is always an eigenvalue for any stochastic matrix \( P \).

\[
Ax = \lambda x \Rightarrow xP = \lambda x.
\]
29. Lecture 27: 12/03/03

29.1. Gambler’s Ruin Problem. If $p = q = \frac{1}{2}$,

\begin{equation}
Q(x) = \frac{x}{x_{max}},
\end{equation}

\begin{equation}
P(x) = 1 - \frac{x}{x_{max}}
\end{equation}

and

\begin{equation}
\lim_{x_{max} \to \infty} P(x) = 1.
\end{equation}

Similarly, one can argue that with probability 1, the party with no capital limitation wins. In fact, as long as $q \leq p$, the $\lim_{x_{max} \to \infty} P(x) = 1$.

**Exercise 12.** Show that

\begin{equation}
\lim_{x_{max} \to \infty} P(x) = 1
\end{equation}

when $q \geq p$.

Aside: Back to the Martingale Convergence Theorem.

\begin{equation}
\sup_n E[|x_n|] < \infty,
\end{equation}

then $x_n \to x$ a.s, where $x_n$ is a Martingale sequence and $x$ is a r.v.

Markov Chain: $x_0, x_1, x_2, \ldots, x_0$ is a r.v. with distribution $P_0: P_0\{x_0 = i\} = P_0(i)$. Define

\begin{equation}
P_{ij} = P\{x_{n+1} = j| x_n = i\}.
\end{equation}

$P = ((P_{ij}))$.

\begin{equation}
P\{x_{n+2} = j|x_n = i\} = \sum_k P\{x_{n+2} = j, x_{n+1} = k|x_n = i\}.
\end{equation}

Use Bayes Identity:

\begin{equation}
P(A \cap B|C) = P(A|B \cap C)P(B|C).
\end{equation}

\begin{equation}
\sum_k P\{x_{n+2} = j \cap x_{n+1} = k| x_n = i \cap x_{n+1} = k\} P\{x_{n+1} = k|x_n = i\}
\end{equation}

\begin{equation}
= \sum_k P\{x_{n+2} = j| x_n = i, x_{n+1} = k\} P\{x_{n+1} = k|x_n = i\}
\end{equation}

\begin{equation}
= \sum_k P\{x_{n+2} = j| x_{n+1} = k\} P\{x_{n+1} = k|x_n = i\}
\end{equation}

\begin{equation}
= \sum_k P_{ik}P_{kj} = (P \times P)_{i,j} = P_{i,j}^2.
\end{equation}
More generally,

\[ P[x_{n+k} = j | x_n = i] = P^n(i, j). \]

\( P^k \) is the \( k \)-step transition probability matrix.

Back to Gambler’s ruin problem: \( x_{\text{max}} = 4 \), \( p = 0.6 \) and \( q = 0.4 \).

\[
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 \\
0.4 & 0 & 0.6 & 0 & 0 \\
0 & 0.4 & 0 & 0.6 & 0 \\
0 & 0 & 0.4 & 0 & 0.6 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
\]

Regression: Also, we know that \( P^n = P^k P^\ell \), where \( \ell + k = n \).

Back to problem:

\[
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 \\
0.4 & 0.24 & 0 & 0.36 & 0 \\
0.16 & 0 & 0.48 & 0 & 0.36 \\
0 & 0.16 & 0 & 0.24 & 0.6 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
\]

\[
\lim_{n \to \infty} P^n = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 \\
0.5846 & 0 & 0 & 0 & 0.4154 \\
0.3077 & 0 & 0 & 0 & 0.6923 \\
0.1231 & 0 & 0 & 0 & 0.8769 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
\]

(Use Spectral Theorem to solve). Note that the last column is

\[
Q(x) = \begin{bmatrix} Q(0) \\ \vdots \\ Q(4) \end{bmatrix}.
\]

In our case,

\[
P_o = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \end{bmatrix}.
\]

In particular,

\[
(P_o P^\infty)_{x_{\text{max}}} = (Q(x))_{x_{\text{max}}}
\]

\[
\lim_{n \to \infty} \{ x_n = x_{\text{max}} | x_o = x \}.
\]
(363) \[ P_n = P(x_n = j) = P_0, P^n = P_1 P^{n-1} = P_2 P^{n-2} = \ldots = P_{n-1} P. \]

But, what is the limit as \( n \to \infty \) of \( P_n \)?

(364) \[ \lim_{n \to \infty} P_n = P_0 P^\infty. \]

\[ \therefore P_n = P_{n-1} P. \]

(365) \[ \lim_{n \to \infty} P_n = ( \lim_{n \to \infty} P_{n-1} ) P. \]

(366)

where \( \pi \) is the corresponding \( \pi \).

Consider:

\[ \begin{array}{ccc}
\text{upstate} & P_{12} & \text{downstate} \\
1 & P_{11} & 0 \\
\text{downstate} & P_{21} & \text{upstate}
\end{array} \]

\[ P = \begin{bmatrix}
p_{11} & p_{12} \\
p_{21} & p_{22}
\end{bmatrix} = \begin{bmatrix}
0.9999 & 0.0001 \\
0.01 & 0.99
\end{bmatrix}. \]

(367)

\[ \lim_{n \to \infty} P^n = \begin{bmatrix}
0.9901 & 0.0099 \\
0.9901 & 0.0099
\end{bmatrix}. \]

Rows are identical: meaning the initial state does not affect the state in the limit as \( n \to \infty \).

\( P(\text{up at } n = \infty) = 0.9901 \). \( P(\text{down at } n = \infty) = 0.0099 \).

(369) \[ P\{x_n = i\} = (P_0 P)_{\text{ith component}}. \]

(370) \[ \pi_n = \begin{bmatrix}
P\{x_n = 1\} \\
P\{x_n = 0\}
\end{bmatrix} = P_0 P^n. \]

Take limits:

(371) \[ \lim_{n \to \infty} \pi_n = \lim_{n \to \infty} P_0 P^n \]
\[(98) \quad \lim_{n \to \infty} \pi_{n-1} = \pi \nabla P.\]

\[(373) \quad \pi = \pi P.\]

\[(374) \quad \therefore \quad \lim_{n \to \infty} P^n = \begin{bmatrix} \pi \\
\vdots \\
\pi \end{bmatrix}.

Suppose that \(P_0 = \pi\).

\[(375) \quad \pi_1 = P_0 \pi = \pi P = \pi.

\[(376) \quad \pi_2 = P_0 P^2 = P_0 \pi P = \pi P = \pi.

\[(377) \quad \vdots

\[(378) \quad \pi_n = \pi.

\therefore \pi is called a stationary distribution.

Finding \(\pi\): Set \(\pi = \pi P\).

\[(379) \quad \begin{bmatrix} \pi_1 \\
\pi_2 \\
\end{bmatrix} = \begin{bmatrix} \pi_1 \\
\pi_2 \\
\end{bmatrix} \begin{bmatrix} P_{11} & P_{12} \\
P_{21} & P_{22} \end{bmatrix}.

\[(380) \quad \pi_1 = P_{11} \pi_1 + P_{21} \pi_2 \quad \therefore \pi_1 (1 - P_{11}) = P_{21} \pi_2.

\[(381) \quad \pi_2 = P_{12} \pi_1 + P_{22} \pi_2 \quad \therefore \pi_1 = \frac{P_{21}}{1 - P_{11}} \pi_2.

\[(382) \quad \pi = \begin{bmatrix} \pi_1 \\
\frac{1 - P_{11}}{P_{21}} \pi_1 \end{bmatrix}

with the proviso that: \(\sum_i \pi_i = 1\). \therefore \pi_1 + \frac{1 - P_{11}}{P_{21}} \pi_1 = 1.

\[(383) \quad \therefore \pi_1 \left(1 + \frac{1 - P_{11}}{P_{21}}\right) = 1.

\[(384) \quad \pi_1 = \frac{P_{21}}{P_{21} - P_{11} + 1}; \quad \pi_2 = 1 - \pi_1.

Sufficient condition for existence of \(\pi\): If every component is non-zero, then \(\pi P = \pi\), \(\sum_i \pi_i = 1\) has a unique non-zero solution called the stationary distribution of the Markov Chain. If

\[(385) \quad P_{ii}^n = P\{x_n = i|x_0 = i\}

and

\[(386) \quad P_{ij}^n = P\{x_n = j|x_0 = i\},\]
then

\[(387)\]

\[\pi_i = \lim_{n \to \infty} P^n_{ii} = \lim_{n \to \infty} P^n_{ji}.\]

A process \(x(t)\) is ergodic if

\[(388)\]

\[E[x(t)] = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x(t) dt \quad \text{a.s.} \]

30.1.1. The Deterministic Case. Let \( x(t) \) be a signal with \( x(f) = 0, \forall |f| > w \), where

\[
x(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt.
\]

![Figure 68.](image)

Define

\[
s(t) \triangleq \sum_{k=-\infty}^{\infty} \delta(t - kT).
\]

Note that:

\[
s(f) = \frac{1}{T} \sum_{k=-\infty}^{\infty} \delta\left(f - \frac{k}{T}\right).
\]

![Figure 69.](image)

Proof:

\[
s(t) = \sum_{k} \delta(t - kT)
\]

\[
= \sum_{k} C_k e^{j\frac{2\pi k}{T}},
\]

where

\[
C_k = \frac{1}{T} \int_{-T/2}^{T/2} s(t)e^{-j\frac{2\pi kt}{T}} dt = \frac{1}{T} \cdot 1 = \frac{1}{T}.
\]
s(t) = \frac{1}{T} \sum_k e^{j \frac{2\pi k t}{T}}.

\Rightarrow s(f) = \frac{1}{T} \sum_k \delta \left(f - \frac{k}{T}\right).

\frac{1}{T} \geq 2W. \text{ Define } x_s(t) = x(t) s(t).

x_s(t) = x(t) \sum_{k=-\infty}^{\infty} \delta(t - kT) = \sum_{k} x(kT) \delta(t - kT).

x_s(f) = \mathcal{F}\{x(t) s(t)\}.

x_s(f) = X(f) \ast S(f) = X(f) \ast \frac{1}{T} \sum_k \delta \left(f - \frac{k}{T}\right).

X_s(f) = \frac{1}{T} \sum_k x \left(f - \frac{k}{T}\right).

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure70.png}
\caption{Figure 70.}
\end{figure}

\frac{1}{T} - W \geq W. \frac{1}{T} \geq 2W \text{ so we have no overlap.}

H(f) = T\text{rect}(2W f).

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure71.png}
\caption{Figure 71.}
\end{figure}

X(f) = H(f) X_s(f).

x(t) = h(t) \ast s_x(t).
\( h(t) = \mathcal{F}^{-1}\{\text{rect}(2Wf)\} = 2WT\text{sinc}(2\pi Wt). \)

\( x(t) = 2WT \sum_k x(kT) \sin\left(\frac{2\pi W(t - kT)}{2\pi W(t - kT)}\right). \)

Pick \( T = \frac{1}{2W} \) or \( \frac{1}{T} = 2W \) (Nyquist Rate).

\( x(t) = \sum_k x(kT) \frac{\sin\left(\frac{2\pi W(t - \frac{kT}{2\pi W})}{2\pi W(t - \frac{kT}{2\pi W})}\right)}{2\pi W(t - \frac{kT}{2\pi W})}. \)

Remark: If \( x(t) \) is an energy signal, i.e., \( \int x^2(t)dt < \infty \), then the above representation of \( x(t) \) says the following:

\[ \lim_{N \to \infty} \int \left| x(t) - \sum_{k=-N}^{N} x(kT) \frac{\sin\left(\frac{2\pi W(t - \frac{kT}{2\pi W})}{2\pi W(t - \frac{kT}{2\pi W})}\right)}{2\pi W(t - \frac{kT}{2\pi W})} \right|^2 dt = 0. \]

30.1.2. The Stochastic Case. A process \( x(t) \) is said to be bandlimited to \( W \), if its power spectral density \( \phi(f) = 0, \forall |f| > W. \)
\[
\phi(f) = \lim_{T \to \infty} \frac{E[|\mathcal{F}[X_T(t)]|^2]}{2T}.
\]

\[
X_T(t) = \begin{cases} 
X(t), & |t| \leq T \\
0, & \text{otherwise}
\end{cases}
\]

From the Wiener-Kinchin relation,
\[
\phi(f) = \mathcal{F}\{\phi(\tau)\},
\]
where \(\phi(\tau)\) is a deterministic function.

\[
\phi(\tau) = \sum_k \phi(kT) \frac{\sin(2\pi W(t - \frac{k}{2W}T))}{2\pi W(t - \frac{k}{2W}T)}.
\]

30.2. **Stochastic Sampling Theorem.** If a WSS random process \(X(t)\) is bandlimited to the bandwidth \(W\), then upon setting \(T = \frac{1}{2W}\) (Nyquist Rate)

\[
X(t) = \sum_k X(kT) \frac{\sin(2\pi W(t - \frac{k}{2W}T))}{2\pi W(t - \frac{k}{2W}T)}.
\]

in the mean square sense, that is with
\[
X_N(T) \equiv \sum_{n=-N}^{N} X(nT) \frac{\sin(2\pi W(t - \frac{k}{2W}T))}{2\pi W(t - \frac{k}{2W}T)}.
\]
Then,

\[
\lim_{N \to \infty} E\{(X(t) - X_N(t))^2\} = 0, \text{ for each } t.
\]

Proof: We want to prove
\[
\lim_{N \to \infty} E\{(X(t) - X_N(t))^2\} = 0
\]
\[
= \lim_{N \to \infty} E\{(X(t) - X_N(t))(X(t) - X_N(t))\}
\]
\[
= \lim_{N \to \infty} \{E[(X(t) - X_N(t))X(t)] - E[(X(t) - X_N(t))X_N(t)]\}
\]
\[
\lim_{N \to \infty} E[(X(t) - X_N(t))X(t)] = E[X^2(t)] - \lim_{N \to \infty} E[X_N(t)X(t)]
\]
\[
= \phi(0) - \sum_{k=-\infty}^{\infty} E[X(t)X(kT)] \frac{\sin(2\pi W(t - \frac{k}{2W}T))}{2\pi W(t - \frac{k}{2W}T)}.
\]
\[
= \phi(0) - \sum_{k=-\infty}^{\infty} \phi(kT - t) \frac{\sin(2\pi W(t - \frac{k}{2W}T))}{2\pi W(t - \frac{k}{2W}T)}.
\]
\[
= \phi(0) - \phi(0) = 0.
\]
\( X_N(t) = \sum_{n=-N}^{N} X(nT) \frac{\sin(2\pi W(t - \frac{k}{2T}))}{2\pi W(t - \frac{k}{2T})} \)

So, \( X_N(t) \) is a weighted sum of \( x(mT) \), so it is enough to show that

\[
\lim_{N \to \infty} E[(X(t) - X_N(t))X(mT)] = 0.
\]

\[
\lim_{N \to \infty} E[(X(t) - X_N(t))X(mT)] = E[X(t)X(mT)] - \lim_{N \to \infty} E[X_N(t)X(mT)]
\]

\[
= \phi(t - mT) - \sum_{n=-\infty}^{\infty} E[X(nT)X(mT)] \frac{\sin(2\pi W(t - \frac{k}{2T}))}{2\pi W(t - \frac{k}{2T})}
\]

\[
= \phi(t - mT) - \sum_{n=-\infty}^{\infty} \phi(nT - mT) \frac{\sin(2\pi W(t - nT))}{2\pi W(t - nT)}
\]

\[
= \phi(t - mT) - \phi(t - mT) = 0.
\]
REFERENCES

