ME (MATH) 577

STOCHASTIC SYSTEMS FOR SCIENCE AND ENGINEERING

Part I-1: SAMPLE SPACE AND EVENT SPACE

Definition 1-1: The sample space Ω is a nonempty set of experimental outcomes (also called sample points) ζ.

Remark 1-1: The sample space Ω can be finite, countably infinite $(\aleph_0)$, or uncountably infinite $(\aleph_1)$. ■

Definition 1-2: Given a sample space Ω, a collection of subsets of Ω is called an event set $\mathcal{E}$ provided that the following three conditions hold:

(i) $\Omega \in \mathcal{E}$; (ii) $E^c \in \mathcal{E} \quad \forall E \in \mathcal{E}$ where $E^c \equiv \Omega - E$; (iii) $\bigcup_{k=1}^{\infty} E_k \in \mathcal{E}$ with $E_k \in \mathcal{E}$

The event space $\mathcal{E}$ is also known as a σ-algebra of the sample space Ω. The pair $\langle \Omega, \mathcal{E} \rangle$ is called a measurable space and the members of $\mathcal{E}$ are called measurable sets in Ω. ■

Remark 1-2: Each event E in the event space $\mathcal{E}$ is a measurable set in the sample space Ω. ■

Remark 1-3: If the property (iii) in Definition 1-2 is restricted to finite union, i.e., $\bigcup_{k=1}^{K} E_k \in \mathcal{E}$ if $E_k \in \mathcal{E}$ and $K \in \mathbb{N}$, the set of positive integers, then $\mathcal{E}$ is known as an algebra of the sample space Ω. Therefore, every σ-algebra is an algebra but every algebra is not a σ-algebra. ■

Remark 1-4: The smallest event space of a sample space Ω is $\{\varnothing, \Omega\}$. ■

Remark 1-5: The largest event space of a sample space Ω is its power set 2^{Ω}, i.e., the collection of all subsets of Ω. Therefore, if Ω is a finite set, any event space $\mathcal{E}$ of the sample space Ω must be a finite set. ■

Definition 1-3: A (nonempty) collection $\Im$ of subsets of a (nonempty) set Ω is called a topology provided that the following three conditions hold:

(i) $\varnothing \in \Im$ and $\Omega \in \Im$

(ii) $\bigcup_{\alpha \in X} E_\alpha \in \Im$ if every $E_\alpha \in \Im$ where X is an index set, finite, countable, or uncountable

(iii) $\bigcap_{k=1}^{K} E_k \in \Im$ if every $E_k \in \Im$ and $K \in \mathbb{N}$

The pair $\langle \Omega, \Im \rangle$ is called a topological space and the members of $\Im$ are called open sets in Ω. ■

Definition 1-4: Given a sample space Ω and a nonempty collection $\mathcal{G}$ of nonempty subsets of Ω, the smallest σ-algebra containing the sets in $\mathcal{G}$ is called the minimal σ-algebra over $\mathcal{G}$ or the σ-algebra generated by $\mathcal{G}$. ■

Remark 1-6: Let E be a non-empty proper subset of Ω. Then, the smallest event space containing E is $\{\varnothing, E, E^c, \Omega\}$ which is an example of the σ-algebra generated by the set E. ■

Definition 1-5: Let $\Omega = R \equiv (-\infty, \infty)$. The minimal σ-algebra over the collection of all right semi-closed intervals in R, i.e., $\{ (a,b] \subset R : a < b \}$, is called the Borel set $\Re$. ■

Remark 1-7: The Borel set $\Re$ contains all open intervals, closed intervals, semi-open intervals, and singleton subsets in R as well as their at most countable unions and intersections. We can similarly define the Borel set $\Re^n$ over the n-dimensional space R^n for any $n \in \mathbb{N}$. We can also define the Borel set $\overline{\Re}^n$ over the n-dimensional space $\overline{R}^n$ for any $n \in \mathbb{N}$ where $\overline{R} = [-\infty, \infty]$ is the extended real line. ■

Remark 1-8: If $\mathcal{F}$ is the collection of all finite disjoint unions of right semi-closed intervals in R, then $\mathcal{F}$ is an algebra but not a σ-algebra because the condition (iii) in Definition 1-2 is not completely satisfied. Note that, in this case, $\mathcal{F}$ does not contain any open interval, closed interval, right semi-open interval, and singleton subset of R. ■

Part I-2: Probability MEASURE AND RANDOM VARIABLES

Definition 1-6: A set function μ on a σ-algebra $\mathcal{E}$ is defined to be finite if $|\mu(A)| < \infty \ \ \forall A \in \mathcal{E}$. ∎

Definition 1-7: A σ-finite measure $\mu : \mathcal{E} \to [0, \infty]$ is a countably additive set function on a measurable space $\langle \Omega, \mathcal{E} \rangle$ under the following conditions:

(i) $\mu(\varnothing) = 0$

(ii) $\mu\left(\bigcup_{k=1}^{\infty} E_k \right) = \sum_{k=1}^{\infty} \mu(E_k)$ provided that $E_k \in \mathcal{E} \ \ \forall k \in N$ and $E_k \cap E_j = \varnothing \ \ \forall k \neq j$.

(iii) There exists a sequence $\{E_k\}$ such that $\bigcup_{k=1}^{\infty} E_k = \Omega$ and $\mu(E_k) < \infty \ \ \forall k \in N$. ∎

Definition 1-8: A measure μ on a measurable space $\langle \Omega, \mathcal{E} \rangle$ is complete if $\mathcal{E}$ contains all subsets of *zero measure* sets, i.e., if $F \in \mathcal{E}$ and $\mu(F) = 0$, then $E \in \mathcal{E} \ \ \forall E \subset F$. ∎

Definition 1-9: The probability measure $P : \mathcal{E} \to [0,1]$ is a countably additive set function on an event space (i.e., σ-algebra) $\mathcal{E}$ of a sample space Ω under the following axioms:

Axiom 1: $P(\Omega) = 1$

Axiom 2: $P\left(\bigcup_{k=1}^{\infty} E_k \right) = \sum_{k=1}^{\infty} P(E_k)$ provided that $E_k \in \mathcal{E} \ \ \forall k \in N$ and $E_k \cap E_j = \varnothing \ \ \forall k \neq j$. ∎

Remark 1-9: The probability measure is finite and translation-variant in contrast to the standard Lebesgue measure that is σ-finite and translation-invariant. ∎

Definition 1-10: Let $\langle \Omega_1, \mathcal{E}_1 \rangle$ and $\langle \Omega_2, \mathcal{E}_2 \rangle$ be two measurable spaces. Then, $g : \Omega_1 \to \Omega_2$ is called a measurable function relative to the σ-algebras $\mathcal{E}_1$ and $\mathcal{E}_2$ if $g^{-1}(A) \in \mathcal{E}_1 \ \ \forall A \in \mathcal{E}_2$. ∎

Remark 1-10: Measurability of $g : \Omega_1 \to \Omega_2$ relative to $\mathcal{E}_1$ and $\mathcal{E}_2$ does ***not*** imply $g(A) \in \mathcal{E}_2 \ \ \forall A \in \mathcal{E}_1$. For example, let $\Omega_1 = \Omega_2$ and let $g : \Omega_1 \to \Omega_2$ be the identity function, i.e., $g(x) = x$. If $\mathcal{E}_2 = \{\varnothing, \Omega_2\}$, then g is a measurable function for any σ-algebra $\mathcal{E}_1$ of Ω_1. Let $\mathcal{E}_1$ contain a nonempty proper subset $A \subset \Omega_1$. Then, $g(A) = A \notin \mathcal{E}_2$. ∎

Definition 1-11: Let $\langle \Omega_1, \Im_1 \rangle$ and $\langle \Omega_2, \Im_2 \rangle$ be two topological spaces. Then, $g : \Omega_1 \to \Omega_2$ is called a continuous function relative to the topologies $\Im_1$ and $\Im_2$ if $g^{-1}(A) \in \Im_1 \ \ \forall A \in \Im_2$. ∎

Remark 1-11: The concept of a measurable function is similar to that of a continuous function in the context of two topological spaces. Continuity of $g : \Omega_1 \to \Omega_2$ relative to the respective topologies $\Im_1$ and $\Im_2$ does ***not*** imply $g(A) \in \Im_2 \ \ \forall A \in \Im_1$. For example, let $\Omega_1 = \Omega_2$ and let $g : \Omega_1 \to \Omega_2$ be the identity function, i.e., $g(x) = x$. If $\Im_2 = \{\varnothing, \Omega_2\}$, then g is a continuous function for any topology $\Im_1$ of Ω_1. Let $\Im_1$ contain a nonempty proper subset $A \subset \Omega_1$. Then, $g(A) = A \notin \Im_2$. ∎

Definition 1-12: Let $\langle \Omega, \mathcal{E} \rangle$ be a measurable space. Then, $g : \Omega \to R^n$, $n \in N$, is called a $\mathcal{E}$-measurable function on $\langle \Omega, \mathcal{E} \rangle$ if g is a measurable function relative to $\mathcal{E}$ and $\Re^n$, i.e., $g^{-1}(A) \in \mathcal{E} \ \ \forall A \in \Re^n$. ∎

Remark 1-12: Let $\wp \subseteq R^k$ be a Borel subset, i.e., $\wp \in \Re^k$ for some $k \in N$. If we use the term *Borel measurable function* mapping $\wp$ into R^k for a σ-algebra $\mathcal{E}$ of $\wp$, then it is implied that $\mathcal{E} \subseteq \Re^k$. ∎

Definition 1-13: The probability space is the triple $\langle \Omega, E, P \rangle$ where Ω is a (finite, countable, or uncountable) sample space and E is the event space corresponding to the measurable space $\{\Omega, E\}$; and $P : E \to [0, 1]$ is a probability measure. ■

Definition 1-14: A E-measurable function $X : \Omega \to R^n$, $n \in N$, on a measurable space $\{\Omega, E\}$ is called a real random vector of dimension n. For $n = 1$, $X : \Omega \to R$ is called a real random variable. ■

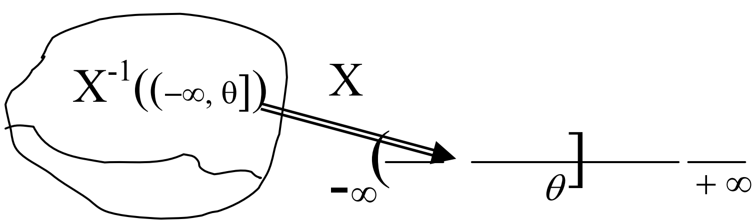
Remark 1-13: The respective probability spaces $\langle \Omega, E, P \rangle$ and $\langle R^n, \Re^n, P_X \rangle$, $n \in N$, are equivalent in the sense that $P_X(A) = P(X^{-1}(A))$ $\forall A \in \Re^n$. For example, letting $n = 1$, we have:

$$P_X((-\infty, a]) = P(\{\zeta \in \Omega : -\infty < X(\zeta) \le a\}) \qquad P_X((-\infty, a)) = P(\{\zeta \in \Omega : -\infty < X(\zeta) < a\})$$
$$P_X((a, b]) = P(\{\zeta \in \Omega : a < X(\zeta) \le b\}) \qquad P_X(\{a\}) = P(\{\zeta \in \Omega : X(\zeta) = a\})$$ ■

Definition 1-15: Let us consider the probability space $\langle R^n, \Re^n, P_X \rangle$, $n \in N$. Instead of using the probability measure P_X that is a set function whose domain is the Borel set $\Re^n$ having the range $[0,1]$, we introduce an equivalent function $F_X : R^n \to [0,1]$ as: $F_X(\theta_1, \theta_2, \cdots) \equiv P_X\left(\bigtimes_{k=1}^{n} (-\infty, \theta_k] \right)$ for every semi-infinite closed cell in R^n.

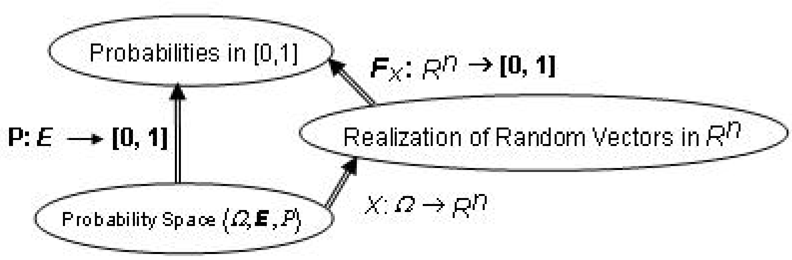
This is the (scalar-valued n-dimensional) joint probability distribution function (PDF) F_X of the random vector $X : \Omega \to R^n$, $n \in N$. Specifically, for $n = 1$, F_X is called the probability distribution function (PDF) of the random variable $X : \Omega \to R$. ■

Remark 1-14: A pictorial representation of probability distribution function is given below.



Remark 1-15: The joint probability distribution function can be expressed as follows:

$$F_X(\theta_1, \theta_2, \cdots) \equiv P_X\left(\bigtimes_{k=1}^{n} (-\infty, \theta_k] \right) = P\left(\bigcap_{k=1}^{n} \{\zeta \in \Omega : -\infty < X_k(\zeta) \le \theta_k\} \right)$$



Remark 1-16: It follows from Definition 1-15 and Remark 1-15 that $F_X : R^n \to [0,1]$ is right-continuous. If we define $F_X(\theta_1, \theta_2, \cdots) \equiv P_X\left(\bigtimes_{k=1}^{n} (-\infty, \theta_k) \right)$, then $F_X : R^n \to [0,1]$ will be left-continuous. ■

Remark 1-17: $F_X(\theta) - F_X(\theta^-) = P(\{\zeta \in \Omega : X(\zeta) = \theta\}) = 0$ if $F_X : R^n \to [0,1]$ is continuous at a point $\theta \in R^n$; otherwise, $(F_X(\theta) - F_X(\theta^-)) > 0$. In this case, we call $(F_X(\theta) - F_X(\theta^-))$ as the probability mass function (PMF) at the point $\theta \in R^n$. ■

Definition 1-16: In a probability space $\langle R^n, \Re^n, P_X \rangle$, $n \in N$, the probability measure P_X is called singular relative to the Lebesgue measure μ if $\exists S \in \Re^n$ such that $P_X(S) \in (0,1]$ and $\mu(S) = 0$. ■

Definition 1-17: Let $\langle R^n, \Re^n, P_X \rangle$, $n \in N$, be the probability space under consideration. The probability measure P_X is absolutely continuous with respect to the Lebesgue measure μ, denoted as $P_X << \mu$, if the condition $\mu(S) = 0 \;\Rightarrow P_X(S) = 0 \;\; \forall S \in \Re^n$ is satisfied.

[Note: *The notion of absolutely continuity of a function is commonly used in a different sense. A function* $\varphi : [a,b] \to R$ *is called absolutely continuous on the interval* $[a,b]$ *if the following condition holds:*

$\forall \varepsilon > 0 \;\; \exists \delta(\varepsilon) > 0$ *such that* $\sum_{k=1}^{n} |\varphi(\tilde{x}_k) - \varphi(x_k)| < \varepsilon$ *for every (finite) collection* $\{\tilde{x}_k, x_k\}$ *of non-overlapping intervals with* $\sum_{k=1}^{n} |\tilde{x}_k) - x_k| < \delta$.

Also note that: (i) every indefinite integral is absolutely continuous; and (ii) Absolute continuity on $[a,b]$ *implies differentiability almost everywhere on* $[a,b]$] ■

Definition 1-18: Let $\langle R^n, \Re^n, P_X \rangle$, $n \in N$, be the probability space under consideration. If $P_X << \mu$, then $\exists$ a non-negative Borel measurable function f such that $P_X(E) = \int_E f \, d\mu \;\; \forall E \in \Re^n$. The function f is uniquely determined $\mu-$almost everywhere on $\Re^n$ and is called the Radon-Nikodym derivative of P_X with respect to μ, denoted as $\frac{dP_X}{d\mu}$. ■

Remark 1-18: For a continuous random vector $X : \Omega \to \Re^n$, $n \in N$, the Radon-Nikodym derivative of P_X with respect to the Lebesgue measure μ can be expressed at a point $\theta \in R^n$ as: $\left.\frac{dP_X}{d\mu}\right|_{\theta \in R^n} = \frac{\partial^n F_X(\theta_1, \cdots, \theta_n)}{\partial\theta_1 \cdots \partial\theta_n}$ and is known as the joint probability density function (pdf) that is denoted as: $f_X(\theta)$ or $f_X(\theta_1, \cdots, \theta_n)$. ■

Remark 1-19: Let two probability measures, $P_0 : \Re^n \to [0,1]$ and $P_1 : \Re^n \to [0,1]$, be defined on a measurable space $\langle R^n, \Re^n \rangle$, $n \in N$, to represent two hypotheses — nominal and faulty conditions. If $P_1 << P_0 << \mu$, then the likelihood ratio of these two hypotheses at a point $\theta \in R^n$ is expressed as: $\left.\frac{dP_1}{dP_0}\right|_{\theta \in R^n} = \left.\frac{dP_1/d\mu}{dP_0/d\mu}\right|_{\theta \in R^n} = \frac{f_1(\theta)}{f_0(\theta)}$. ■

Remark 1-20: A probability measure P could be neither singular nor absolutely continuous. Using Lebesgue decomposition, one can claim that $\exists$ a singular measure P_s and an absolutely continuous measure P_a and $\alpha \in [0,1]$ such that $P = \alpha P_s + (1-\alpha) P_a$.

Definition 1-19: A set of random variables, $X_1, X_2, \cdots X_n$, having a given distribution function $F(\theta)$, $\theta \in R^n$, $n \in N$, is called a realization of $F(\bullet)$. ■

Remark 1-21: An n-dimensional (scalar-valued) probability distribution function $F(\theta), \theta \in R^n$, defines a probability measure P_X on the measurable space $\langle R^n, \Re^n \rangle$ where $X : R^n \to R^n$ is a Borel-measurable function defined as:

$X_j(\varphi)=\theta_j, \varphi \in R^n$. Then, $X \equiv (X_1, X_2, \cdots X_n)$ as a random vector on the probability space $\langle R^n, \Re^n, P\rangle$ will have $F_X(\bullet)$ as the joint distribution function because:

$$F_X(\theta_1,\theta_2;\cdots\theta_n)=F_X(\theta)=P\left(\left\{\varphi\in R^n : X_1(\varphi)\le\theta_1;\ X_2(\varphi)\le\theta_2;\cdots\ X_n(\varphi)\le\theta_n\right\}\right) \qquad \blacksquare$$

Definition 1-20: Let $\{E_k\}$ be a sequence of events on a probability space $\langle \Omega, E, P\rangle$. We define superior and inferior limits as follows: $\limsup_{n\to\infty} E_n \equiv \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} E_k$ and $\liminf_{n\to\infty} E_n \equiv \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} E_k$

If the two limits coincide, then we say that $\{E_k\}$ is a convergent sequence of events and the limit set is defined as:

$$E=\lim E_n=\liminf E_n=\limsup E_n \qquad \blacksquare$$

Remark 1-22: The superior limit is the set of points that occur in infinitely many $E_k^{'}s$, while the inferior limit is the set of points that occur in all but finitely many $E_k^{'}s$ such that $\liminf E_n \subseteq \limsup E_n$. The limit concept is similar to that for a sequence $\{x_k\}$of real numbers where $\liminf_{n\to\infty} x_n \equiv \sup_n \inf_{k\ge n} x_k$ and $\limsup_{n\to\infty} x_n \equiv \inf_n \sup_{k\ge n} x_k$. In general, $\liminf x_n \le \limsup x_n$; and $\{x_k\}$ converges to the limit x if $x \equiv \lim x_n = \liminf x_n = \limsup x_n$. $\blacksquare$

Remark 1-23: Let $\{X_k\}$ be a sequence of real random variables and $a \in R$. Both $\liminf X_n$ and $\limsup X_n$ are random variables provided that they are finite at every $\zeta \in \Omega$ as explained below:

$$\{\zeta\in\Omega : \liminf X_n(\zeta)\le a\}=\left\{\zeta\in\Omega : \sup_n \inf_{k\ge n} X_n(\zeta)\le a\right\}=\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}\{\zeta\in\Omega :\ X_k(\zeta)\le a\}$$

$$\{\zeta\in\Omega : \limsup X_n(\zeta)\le a\}=\left\{\zeta\in\Omega : \inf_n \sup_{k\ge n} X_n(\zeta)\le a\right\}=\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}\{\zeta\in\Omega :\ X_k(\zeta)\le a\}$$

Note that every set in the form $\{\zeta\in\Omega : \limsup X_n(\zeta)\le a\}$ or $\{\zeta\in\Omega : \liminf X_n(\zeta)\le a\}$ is also an event by the countable union and countable intersection properties of a σ - algebra. Therefore, if $\{X_k\}$ be a sequence of random variables converging to a limit X such that $|X(\zeta)|<\infty\ \forall\zeta\in\Omega$, then X is a random variable. However, if we modify the range of a random variable to extended real line, then $\liminf X_n$ and $\limsup E_k$ are always random variables. $\blacksquare$

Theorem 1-1 (Continuity of Probability Measure): Every probability measure is *sequentially continuous* , i.e., if $\{E_k\}$ is a convergent sequence of events on a probability space $\langle \Omega, E, P\rangle$, then $\lim_{k\to\infty} P(E_k)=P\left(\lim_{k\to\infty} E_k\right)$.

Proof: If $\{E_k\}$ decreases to the empty set $\varnothing$, then the assertion is trivially true. If $\{E_k\}$ monotonically decreases to a non-empty set E, then $P(E_k)=P((E_k \setminus E)\cup E)=P(E_k\setminus E)+P(E)$. Since $\{E_k\setminus E\}$ decreases to $\varnothing$, $P(E_k)\to P(E)$.

Next consider $\{E_k\}$ to be convergent but not necessarily monotone. Let $F_k \equiv \bigcap_{j\ge k} E_j$ and $G_k \equiv \bigcup_{j\ge k} E_j$ implying that $F_k \subseteq E_k \subseteq G_k$ and each of $\{F_k\}$ and $\{G_k\}$ is monotone. Therefore,

$$\left.\begin{aligned} P\left(\lim_{k\to\infty} F_k\right)&=\lim_{k\to\infty} P(F_k)\le \lim_{k\to\infty} P(E_k) \\ P\left(\lim_{k\to\infty} G_k\right)&=\lim_{k\to\infty} P(G_k)\ge \lim_{k\to\infty} P(E_k)\end{aligned}\right\}$$

Since $\lim_{k\to\infty} F_k = \lim_{k\to\infty} E_k = \lim_{k\to\infty} G_k$ by definition, it follows that

$P\left(\lim_{k\to\infty} E_k\right)\le \lim_{k\to\infty} P(E_k)$ and $\lim_{k\to\infty} P(E_k)\le P\left(\lim_{k\to\infty} E_k\right)$. This proves the theorem. $\blacksquare$

Corrolary to Theorem 1-1: Let $\{E_k\}$ be a monotonically increasing sequence of events on a probability space $\langle \Omega, E, P \rangle$. Then, $\lim_{n\to\infty} P[E_n] = P\left[\bigcup_{k=1}^{\infty} E_k\right]$.

Proof: Given $E_k \subseteq E_{k+1}$, let $D_k \equiv E_{k+1} \setminus E_k$ with $D_0 \equiv \varnothing$. Then, $D_i \cap D_j = \varnothing \; \forall i \neq j$ and $E_n = \bigcup_{k=1}^{n} D_k$. Then, $\bigcup_{k=1}^{n} D_k = \bigcup_{k=1}^{n} E_k = E_n$. Therefore, $\lim_{n\to\infty} P[E_n] = \lim_{n\to\infty} \sum_{k=1}^{n} P[D_k] \equiv \sum_{k=1}^{\infty} P[D_k] = P\left[\bigcup_{k=1}^{\infty} D_k\right] = P\left[\bigcup_{k=1}^{\infty} E_k\right]$ ■

Borel-Cantelli Lemma: Let $\{E_k\}$ be an arbitrary sequence of events on a probability space $\langle \Omega, E, P \rangle$. Then, $\sum_{k=1}^{\infty} P(E_k) < \infty$ implies that $P(\limsup E_k) = 0$.

Proof: It follows from Theorem 1-1 that $P(\limsup E_n) = P\left[\lim_{n\to\infty} \bigcup_{k\geq n} E_k\right] \leq \lim_{n\to\infty} \sum_{k\geq n} P[E_k]$. Since $\sum_{k=1}^{\infty} P(E_k) < \infty$, the tail end of the series must sum to zero. Hence, $\lim_{n\to\infty} \sum_{k\geq n} P[E_k] = 0$. This proves the lemma. ■

Definition 1-21: Let A be an event on a probability space $\langle \Omega, E, P \rangle$. We define the indicator function as follows:

$$I_A(\zeta) \equiv \begin{cases} 1 & \text{if } \zeta \in A \\ 0 & \text{if } \zeta \notin A \end{cases} \quad \text{on } \Omega.$$

Given events $A_1, \cdots, A_n$ and $\alpha_1, \cdots, \alpha_n \in R$, a function $S \equiv \sum_{k=1}^{n} \alpha_k I_{A_k}$ on Ω is called a simple random variable. ■

Theorem 1-2: Every random variable is the limit of a sequence of simple random variables. Every non-negative random variable is the limit of an increasing sequence of simple random variables.

Proof: We define a sequence of simple random variables as follows:

$$X_n(\zeta) \equiv \begin{cases} -2^n & \text{if} \quad X(\zeta) \leq -2^n \\ k/2^n & \text{if} \quad X(\zeta) \in [k/2^n, (k+1)/2^n] \text{ for } -2^{2n} \leq k \leq 2^{2n} - 1 \\ 2^n & \text{if} \quad X(\zeta) \geq 2^n \end{cases}.$$

For a fixed ζ and $n \geq \log_2 |X_n(\zeta)|$, we have $\sup_{k\geq n} |X(\zeta) - X_k(\zeta)| \overset{n\to\infty}{\to} 0$ so that $\{X_n\}$ converges X at every ζ. If X is non-negative, then $\{X_n\}$, as defined above, is an increasing sequence of non-negative simple random variables. ■

PART I-3: EXPECTATION OF RANDOM VARIABLES AND RANDOM VECTORS

Let $\langle \Omega, E, P \rangle$ be a probability space. Given the distribution function $F_X : R \to [0,1]$ of a random variable X, the expectation of X is often defined in terms of the Riemann-Stieltjes integral as: $E[X] = \int_{-\infty}^{\infty} \theta \, dF_X(\theta)$ provided that the integral is well defined (i.e., at least one of the two integrals $\int_0^{\infty} \theta \, dF_X(\theta)$ and $-\int_{-\infty}^{0} \theta \, dF_X(\theta)$ is less than ∞). Furthermore, if the random variable X has a defined density function f_X (i.e., if the probability measure P_X is absolutely continuous relative to the Lebesgue measure), then the expectation of X is often defined in terms of the Riemann integral as: $E[X] = \int_{-\infty}^{\infty} \theta \, f_X(\theta) d\theta$. We will provide an alternative definition of expectation in terms of the Lebesgue-Stieltjes integral that is not only more rigorous but also clarifies the relationship between the expectation operator $E[\bullet]$ and the probability measure P.

We first assume that X is a simple random variable, i.e., there exist events $A_1, \cdots, A_n$ and $\alpha_1, \cdots, \alpha_n \in R$ such that $X(\zeta) = \sum_{k=1}^{n} \alpha_k I_{A_k}(\zeta) \quad \forall \zeta \in \Omega$ and I_A is an indicator function, implying that $I_A(\zeta) = \begin{cases} 1 & if\ \zeta \in A \\ 0 & otherwise \end{cases}$. Thus defined, $E[X]$ must satisfy the following four properties:

- $E[X+Y] = E[X] + E[Y]$: Additivity
- $E[cX] = cE[X] \quad \forall c \in R$: Homogeneity
- $X \geq Y \Rightarrow E[X] \geq E[Y]$: Order preservation
- If a sequence $\{X_k\}$ of simple random variables converges to X, then $\lim_{k\to\infty} E[X_k] = E[X]$.

Next let X be a nonnegative (not necessarily simple) random variable and let $\{X_k\}$ be an increasing sequence of nonnegative simple random variables converging to X. Since $\{E[X_k]\}$ is an increasing sequence of nonnegative real numbers, $\lim_{k\to\infty} E[X_k] = \sup_k E[X_k]$ always exists (but may be infinite). To unambiguously define $\lim_{k\to\infty} E[X_k] = E[X]$ for nonnegative random variables, we make use of the following proposition.

Proposition 1-1: Let $\{X_k\}$ and $\{Y_k\}$ be two increasing sequences of nonnegative random variables converging to the same limit X. Then, $\lim_{k\to\infty} E[X_k] = \lim_{k\to\infty} E[Y_k]$.

Proof: HW Exercise that might be discussed in the recitation class. ■

Now let us consider the general case of random variables by removing the restriction of nonnegativity. We express the random variable X as: $X(\zeta) = X^+(\zeta) - X^-(\zeta) \quad \forall \zeta \in \Omega$ where both X^+ and X^- are nonnegative. We define: $E[X] = E[X^+] - E[X^-]$ provided that the right hand side is not of the form $\infty - \infty$. In this way, $E[X]$ satisfies the three postulated properties of additivity, homogeneity, and order preservation. The uniqueness of the limit: $\lim_{k\to\infty} E[X_k] = E[X]$ is established by Proposition 1-1. Now, we introduce the following definition of $E[X]$.

Definition 1-22: Let $\langle \Omega, E, P \rangle$ be a probability space and $X : \Omega \to R$ be a random variable defined on the probability space $\langle R, \Re, P_X \rangle$. The expectation of X is defined in terms of one of the following two Lebesgue-Stieltjes integrals:

$$E[X] \equiv \int_\Omega X(\zeta) P(d\zeta) \text{ or } E[X] = \int_R \theta P_X(d\theta) \text{ denoted as: } \int_\Omega X\, dP \text{ and } \int_R \theta\, dP_X \text{, respectively.}$$ ■

Definition 1-23: A random variable X is said to be integrable (with respect to a measure P) if $E[|X|] < \infty$. ■

Remark 1-24: If $g : R \to R$ be a Borel-measurable function, then

$$E[g(X)] \equiv \int_\Omega g(X)\, dP \text{ or equivalently, } E[g(X)] = \int_R g(\theta)\, dP_X$$ ■

Next we present the following results on sequences of integrable random variables, which are counterparts of standard results on sequences of integrable functions in real analysis.

Proposition 1-2 (Monotone Convergence): Let $\{X_k\}$ be an increasing sequence of nonnegative random variables converging to X. Then,

$$E[X] = \lim_{k\to\infty} E[X_k]$$ ■

Proposition 1-3 (Fatou's Lemma): Let $\{X_k\}$ be a sequence of nonnegative random variables such that $\exists$ an integrable random variable X having the property: $X_k(\zeta) \geq X(\zeta)$ for all k and all ζ. Then,

$$\liminf_{k\to\infty} E[X_k] \geq E\left[\liminf_{k\to\infty} X_k\right] \qquad \blacksquare$$

Proposition 1-4 (Dominated Convergence): Let $\{X_k\}$ be a sequence of random variables converging to X such that $\exists$ an integrable nonnegative random variable Y having the property: $E[Y] < \infty$ and $|X_k(\zeta)| \leq Y(\zeta)$ for all k and all ζ. Then,

$$E[|X|] < \infty\,;\ \lim_{k\to\infty} E[|X_k - X|] = 0\ ;\text{ and } \lim_{k\to\infty} E[X_k] = E[X] \qquad \blacksquare$$

It is relatively straight-forward to extend the above concept of expectation from random variables to random vectors. Let a random vector $X \equiv [X_1, X_2, \cdots, X_n]^T : \Omega \to R^n$ be a random vector defined on the probability space $\langle R^n, \Re^n, P_X \rangle$ with the joint distribution function $F_X : R^n \to [0,1]$ and let $g : R^n \to R^m$ be a Borel-measurable function. Then,

$$E[X] \equiv \int_{\Omega} X\, dP = \int_{R^n} \theta\, dP_X \text{ and } E[g(X)] \equiv \int_{\Omega} g(X) dP = \int_{R^n} g(\theta) dP_X$$

Remark 1-25: If we define $Y = g(X)$ as another random vector $Y : \Omega \to R^m$ defined on the probability space $\langle R^m, \Re^m, P_Y \rangle$, then $E[g(X)] \equiv \int_{\Omega} g(X) dP = \int_{R^m} \varphi\, dP_Y$. Also, $E[X] \in R^n$ expressed as: $E[X] = [E[X_1] \cdots E[X_n]]^T$ and $E[g(X)] \in R^m$ expressed as: $E[Y] = [E[Y_1] \cdots E[Y_m]]^T$. One can write the expectations of the individual components of a random vector in terms of its joint distribution function as:

$$E[X_k] \equiv \int_{R^n} \theta_k\, dF_X(\theta_1, \cdots, \theta_n),\ k = 1, \cdots, n\,;\ E[g(X)_\ell] \equiv \int_{R^n} g(\theta_1, \cdots, \theta_n)_\ell\, dF_X(\theta_1, \cdots, \theta_n),\ \ell = 1, \cdots, m\,;$$

$$\text{or } E[g(X)_\ell] = E[Y_\ell] \equiv \int_{R^m} \varphi_\ell\, dF_Y(\varphi_1, \cdots, \varphi_\ell),\ \ell = 1, \cdots, m \qquad \blacksquare$$

Definition 1-24: Let a random vector $X : \Omega \to R^n$ have a distribution function F_X. Then, the characteristic function of X is defined as:

$$\Phi_X(\xi) \equiv \int_{R^n} \exp\left(i 2\pi \xi^T \theta\right) dF_x(\theta) = \int_{R^n} \exp\left(i 2\pi \sum_{k=1}^n \xi_k \theta_k\right) dF_x(\theta_1, \cdots, \theta_n)$$

Remark 1-26: If X has the density function f_X, the characteristic function can also be expressed as:

$$\Phi_X(\xi) \equiv \int_{R^n} \exp\left(i 2\pi \xi^T \theta\right) f_X(\theta) d\theta = \int_{R^n} \exp\left(i 2\pi \sum_{k=1}^n \xi_k \theta_k\right) f_X(\theta_1, \cdots, \theta_n) d\theta_1 \cdots d\theta_n$$

Notice that the characteristic function is identical to the Fourier transform of the density function for negative frequency, i.e., $\Phi_X(\xi) = \hat{f}_X(-\xi)$. For existence of $\hat{f}_X$ is guaranteed by the fact that $f_X \in L_1(R^n)$, i.e., f_X is absolute integrable over R^n. Therefore, the density function can be generated from the characteristic function by the inversion formula as follows:

$$f_X(\theta) \equiv \int_{R^n} \exp\left(-i 2\pi \xi^T \theta\right) \Phi_X(\xi) d\xi = \int_{R^n} \exp\left(-i 2\pi \sum_{k=1}^n \xi_k \theta_k\right) \Phi_X(\xi_1, \cdots, \xi_n) d\xi_1 \cdots d\xi_n \qquad \blacksquare$$

Definition 1-25: Let a random vector $X : \Omega \to R^n$ have a distribution function F_X. Then, the moment generating function of X is defined as a function of a complex vector $z \in \boldsymbol{C}^n$:

$$\Theta_X(z) \equiv E\left[\exp\left(2\pi z^T X\right)\right] = \int_{R^n} \exp\left(2\pi z^T \theta\right) dF_x(\theta) = \int_{R^n} \exp\left(2\pi \sum_{k=1}^n z_k \theta_k\right) dF_x(\theta_1, \cdots, \theta_n)$$

provided that the random vector $\exp\left(2\pi z^T X\right)$ is integrable. ■

Remark 1-27: The moment generating function is analogous to Laplace transform as the characteristic function is analogous to Fourier transform. Setting the real part of z approaches zero, then $\Theta_X(z)$ approaches $\Phi_X(\operatorname{Im} z)$. Notice that, like Laplace transform, the region of convergence of the moment generating function needs to be specified. ■

Remark 1-28: For a random variable $X:\Omega \to R$, the moment generating function $\Theta_X(z)$ is given by:

$$\Theta_X(z) = E[\exp(z\,X)] = I + z\,E[X] + \frac{z^2}{2!}E\left[X^2\right] + \frac{z^3}{3!}E\left[X^3\right] + \cdots$$

If $\Theta_X(z)$ is analytic in the neighborhood of $z=0$, we have $E\left[X^k\right] = \left.\frac{d^k}{dz^k}\Theta_X(z)\right|_{z=0}$. ■

Proposition 1-5 (Generalized Markov Inequality): Let $\langle \Omega, E, P\rangle$ be a probability space and let $X:\Omega \to R$ be a random variable with the distribution function F_X. Then, $\forall r \in (0,\infty)$ and $\forall \delta \in (0,\infty)$

$$P\left[\left\{\zeta \in \Omega : \left|X(\zeta)\right| \ge \delta\right\}\right] \le \frac{1}{\delta^r}\int_\Omega |X|^r\,dP$$

Proof: $\int_\Omega |X|^r\,dP \ge \int_{\{\zeta\in\Omega:|X(\zeta)|\ge\delta\}} |X|^r\,dP \ge \delta^r P\left[\left\{\zeta \in \Omega : \left|X(\zeta)\right| \ge \delta\right\}\right] \Rightarrow P\left[\left\{\zeta \in \Omega : \left|X(\zeta)\right| \ge \delta\right\}\right] \le \frac{1}{\delta^r}\int_\Omega |X|^r\,dP$ ■

Corollary 1 to Proposition 1-5 (Chebyshev Inequality): Let $\mu_X < \infty$ and $\sigma_X^2 < \infty$ are the mean and variance of a random variable X. Then, $\forall \delta \in (0,\infty)$, $P\left[\left\{\zeta \in \Omega : \left|X(\zeta) - \mu_X\right| \ge \delta\right\}\right] \le \left(\frac{\sigma_X}{\delta}\right)^2$.

Proof: The result follows by setting $r = 2$ in Proposition 1-5. ■

Corollary 2 to Proposition 1-5: Let $\mu_X < \infty$ be the mean of a nonnegative random variable X. Then,

$$\forall \delta \in (0,\infty),\ P\left[\left\{\zeta \in \Omega : X(\zeta) \ge \delta\right\}\right] \le \frac{\mu_X}{\delta}.$$

Proof: $\mu_X \equiv \int_\Omega X\,dP \ge \int_{\{X(\zeta)\ge\delta\}} X\,dP \ge \delta \int_{\{X(\zeta)\ge\delta\}} dP = \delta\,P\left[\left\{\zeta \in \Omega : X(\zeta) \ge \delta\right\}\right]$

$\Rightarrow P\left[\left\{\zeta \in \Omega : X(\zeta) \ge \delta\right\}\right] \le \frac{\mu_X}{\delta}$ ■

Part I-4: Convergence of Random Sequences

We consider several modes of convergence of random sequences where the measure is finite. However, for comparison purposes, we also consider the convergence under the σ-finite Lebesgue measure. In general, we consider sequences of measurable functions from the measurable space $(\Omega, \mathbf{E})$ to the Borel-measurable space $(R, \Re)$. With no loss of generality, we assume that $\Omega = R$ and $\mathbf{E} = \Re$ implying that the random variables mapping R into R are Borel-measurable. We consider two measure spaces $(R, \Re, \mu)$ and $(R, \Re, P)$ where the measures P and μ denote the (finite) probability measure and (σ-finite) Lebesgue measure, respectively. Furthermore, we specify $P(R) = 1$ for compatibility with the probability measure. The measure space $(R, \Re, P)$ is more restricted than the measure space $(R, \Re, \mu)$.

Definition 1-25: Let $\{g_k\}$ be a sequence of measurable functions on $(R, \Re, \mu)$ converging to a measurable function g on $(R, \Re, \mu)$ in the following modes:

Uniform convergence: $\forall \varepsilon > 0\ \exists\, n(\varepsilon) \in \mathbf{N}$ such that $\left|g_k(t) - g(t)\right| < \varepsilon\ \ \forall k \ge n\ \ \forall t \in R$

Convergence at a given point $t \in R$: $\forall \varepsilon > 0\ \exists\, n_t(\varepsilon) \in \mathbf{N}$ such that $\left|g_k(t) - g(t)\right| < \varepsilon\ \ \forall k \ge n_t$

Pointwise convergence (also called sure convergence) if the sequence converges at every $t \in R$.

Definition 1-26: Let $\{g_k\}$ be a sequence of measurable functions on $(R,\Re,\mu)$ to a measurable function g on $(R,\Re,\mu)$ in the following modes:

Uniform Cauchy convergence: $\forall \varepsilon > 0\ \exists\, n(\varepsilon) \in \mathbf{N}$ such that $\|g_k(t) - g_\ell(t)\| < \varepsilon\ \ \forall k,\ell \ge n\ \ \forall t \in R$

Cauchy convergence at a given point $t \in R$: $\forall \varepsilon > 0\ \exists\, n_t(\varepsilon) \in \mathbf{N}$ such that $\|g_k(t) - g_l(t)\| < \varepsilon\ \ \forall k,\ell \ge n_t$

Pointwise Cauchy convergence (also called sure Cauchy convergence) if the sequence converges in the Cauchy sense at every $t \in R$.

Almost everywhere (a.e) Cauchy convergence if the sequence converges in the Cauchy sense at every $t \in E$ where $E \in \Re$ and $\mu(R \setminus E) = 0$. Similarly, almost sure (a.s.) Cauchy convergence if the sequence converges in the Cauchy sense at every $t \in E$ where $E \in \Re$ and $P(R \setminus E) = 0$.

Remark 1-27: The above definitions hold for both μ and P. ■

Remark 1-28: Concvergence to a point always implies Cauchy convergence and if the converse is true for all Cauchy sequences, then the space to which these Cauchy sequences belong is called complete. ■

Remark 1-29: Uniform convergence $\Rightarrow$ Sure Convergence $\Rightarrow$ Almost sure convergence. ■

Definition 1-27: Let $r \in [1,\infty)$. Then a measurable function h on $(R,\Re,\mu)$ belongs to the space $L_r(\mu)$ if $\int_R |h(t)|^r\, d\mu < \infty$. The norm is defined as: $\|h\|_{L_r} \equiv \left(\int_R |h(t)|^r\, d\mu\right)^{1/r}$. Similar definitions hold for $L_r(P)$. ■

Remark 1-30: The spaces $L_r(\mu)$ and $L_r(P)$ are complete, i.e., every Cauchy sequence converges in these spaces. ■

Definition 1-28: Let $r \in [1,\infty)$. Then a sequence $\{g_k\}$ of measurable functions in $L_r(\mu)$ converge in $L_r(\mu)$ to a measurable function g in $L_r(\mu)$ if $\forall \varepsilon > 0\ \exists\, n(\varepsilon) \in \mathbf{N}$ such that $\|g_k(t) - g(t)\|_{L_r} < \varepsilon\ \ \forall k \ge n$. Similarly, Cauchy convergence of $\{g_k\}$ is defined as: if $\forall \varepsilon > 0\ \exists\, n(\varepsilon) \in \mathbf{N}$ such that $\|g_k(t) - g_\ell(t)\|_{L_r} < \varepsilon\ \ \forall k,\ell \ge n$. ■

Remark 1-31: Definition 1-28 holds for both μ and P. ■

Remark 1-32: In general, uniform convergence does not imply $L_r(\mu)$ convergence and vice versa. We cite an example to show that uniform convergence does not imply $L_r(\mu)$: Let $g_k = k^{-1/r}\chi_{[0,k]}$. The sequence $\{g_k\}$ converges uniformly to the function 0 but it does not converge to 0 in $L_r(\mu)$. ■

Remark 1-33: In the example of Remark 1-32, sequence $\{g_k\}$ does converge to 0 in $L_r(P)$. Uniform convergence implies convergence in $L_r(P)$ but the converse is not true. ■

Theorem 1-3: Let $P(R) = 1$ and $\{g_k\}$ be a sequence in $L_r(P)$ that converges uniformly on R to a measurable function g. Then, $g \in L_r(P)$ and $\{g_k\}$ converges in $L_r(P)$ to g.

Proof: Uniform convergence implies that $\forall \varepsilon > 0\ \exists\, n(\varepsilon) \in \mathbf{N}$ s.t. $|g_k(t) - g(t)| < \varepsilon\ \ \forall k \ge n\ \ \forall t \in R$. For any $k \ge n$, $\int_R |g_k(t) - g(t)|^r\, dP \le \int_R \varepsilon^r\, dP = \varepsilon^r \le \infty \Rightarrow \|g_k - g\|_{L_r} < \varepsilon < \infty$. Since $(g_k - g) \in L_r(P)$ and $g_k \in L_r(P)$, we have $g \in L_r(P)$ and $\{g_k\}$ converges in $L_r(P)$ to g. ■

Theorem 1-4: Let $\{g_k\}$ be a sequence in $L_r(\mu)$ that converges almost everywhere on R to a measurable function g. Then, if $\exists\, h \in L_r(\mu)$ s.t. $|g_k| \le h$ a.e. on R, then $g \in L_r(\mu)$ and $\{g_k\}$ converges in $L_r(\mu)$ to g.

Proof: Given $|g_k| \le h$ a.e. on R, we have $|g_k(t) - g(t)|^r \le |2h(t)|^r$. Since $|g_k(t) - g(t)|^r \to 0$ a.e. on R, and $|h|^r \in L_1(\mu)$ (because $h \in L_r(\mu)$), we obtain the following result by applying the dominated convergence theorem:

$$\lim_{k\to\infty} \int_R |g_k(t) - g(t)|^r \, d\mu = \int_R \lim_{k\to\infty} |g_k(t) - g(t)|^r \, d\mu = 0 \text{ implies that } g \in L_r(\mu). \qquad \blacksquare$$

Corollary to Theorem 1-4: : Let $\{g_k\}$ be a sequence in $L_r(P)$ that converges almost everywhere on R to a measurable function g. Then, if $\exists$ a constant $C \in [0,\infty)$ s.t. $|g_k| \le C$ a.e. on R, then $g \in L_r(P)$ and $\{g_k\}$ converges in $L_r(P)$ to g.

Proof: Since $P(R) = 1 < \infty$, $C \in L_r(P)$ and the result follows from the theorem. ■

Definition 1-29: A sequence $\{g_k\}$ of measurable functions converges in measure to a measurable function g if $\forall \varepsilon > 0 \;\; \lim_{k\to\infty} \mu(\{t \in R : |g_k(t) - g(t)| \ge \varepsilon\}) = 0$. ■

Remark 1-34: Compare the convergence in measure with almost everywhere convergence that has been defined as:

Convergence almost everywhere: $\forall \varepsilon > 0 \quad P\left[\left\{t \in R : \lim_{k\to\infty} |g_k(t) - g(t)| \ge \varepsilon\right\}\right] = 0$

Convergence in probability: $\forall \varepsilon > 0 \quad \lim_{k\to\infty} P[\{t \in R : |g_k(t) - g(t)| \ge \varepsilon\}] = 0$

The former considers convergence of a sequence of events while the latter considers a sequence of real numbers. ■

Remark 1-35: Uniform convergence $\Rightarrow$ Convergence in measure regardless of whether the measure is finite or not. But this does not hold for pointwise (and hence almost everywhere) convergence for infinite measure. For example, if $g_k = \chi_{[k,k+1]}$, then the sequence $\{g_k\}$ converges pointwise to 0 but it does not converge to 0 in (infinite) measure. ■

Theorem 1-5: Convergence in $L_r(\mu)$ $\Rightarrow$ convergence in measure μ.

Proof: For any $\varepsilon > 0$ let us define $E_k^\varepsilon \equiv \{t \in R : |g_k(t) - g(t)| \ge \varepsilon\}$. Then,

$$\forall \varepsilon > 0 \quad \int_R |g_k(t) - g(t)|^r \, d\mu \ge \int_{E_k^\varepsilon} |g_k(t) - g(t)|^r \, d\mu \ge \varepsilon^r \int_{E_k^\varepsilon} d\mu = \varepsilon^r \mu\left[\{t \in R : |g_k(t) - g(t)| \ge \varepsilon\}\right].$$

Since convergence in $L_r(\mu)$ is equivalent to having $\lim_{k\to\infty} \int_R |g_k(t) - g(t)|^r \, d\mu = 0$, we conclude that $\forall \varepsilon > 0$ $\lim_{k\to\infty} \mu[\{t \in R : |g_k(t) - g(t)| \ge \varepsilon\}] = 0$ implying convergence in measure μ. ■

Remark 1-36: Convergence in measure μ does not imply convergence in $L_r(\mu)$. For example, let $g_k = k\,\chi_{[1/k,\,2/k]}$. Then, $\{g_k\}$ converges to 0 in measure μ but does not converge to 0 in $L_r(\mu)$. ■

Definition 1-30: A sequence $\{g_k\}$ of measurable functions on $(R, \Re, \mu)$ converges almost uniformly to a measurable function g if $\forall \varepsilon > 0 \;\; \exists \;\; E_\varepsilon \in \Re$ with $\mu(E_\varepsilon) < \varepsilon$ s.t. $\{g_k\}$ converges uniformly to g on $R \setminus E_\varepsilon$. The sequence $\{g_k\}$ converges almost uniformly in the Cauchy sense if $\forall \varepsilon > 0 \;\; \exists \;\; E_\varepsilon \in \Re$ with $\mu(E_\varepsilon) < \varepsilon$ s.t. $\{g_k\}$ converges uniformly on $R \setminus E_\varepsilon$ in Cauchy sense. ■

Theorem 1-6: Almost uniform convergence $\Rightarrow$ Almost uniform Cauchy convergence $\Rightarrow$ Convergence almost everywhere.

Proof: The first part of the theorem "Almost uniform convergence $\Rightarrow$ Almost uniform Cauchy convergence" is obvious. To prove the second part, we proceed as follows.

Let a sequence $\{g_k\}$ of measurable functions on $(R, \Re, \mu)$ converge almost uniformly to a measurable function . For $k \in \mathbf{N}$, let $E_k \in \Re$ be such that $\mu(E_k) < 2^{-k}$ and $\{g_n\}$ converges uniformly to g on $R \setminus E_k$. Let $F_j \equiv \bigcup_{j=k}^{\infty} E_j$ implying that $\mu(F_k) < 2^{-k+1}$. Note that $\{g_n\}$ converges uniformly on $R \setminus F_k$ because $E_k \subseteq F_k \Rightarrow R \setminus F_k \subseteq R \setminus E_k$. Let us define functions h_k as:

$$h_k(t) \equiv \begin{cases} \lim_{n \to \infty} g_n(t) & \text{if } t \notin F_k \\ 0 & \text{if } t \in F_k \end{cases}$$

We observe that $\{F_k\}$ is monotonically decreasing and $\mu(F) = 0$ where $F \equiv \bigcap_{k=n}^{\infty} F_k$. If $\ell \le k$, then $h_\ell(t) = h_k(t)$ $\forall\, t \notin F_\ell$. Therefore, $\{h_k\}$ converges to a measurable function on R defined as: $g(t) = h_k(t) \equiv \lim_{n \to \infty} g_n(t)$ *if* $t \notin F_k$.

Hence, $\{g_k\}$ converges to g on $R \setminus F$ implying that $\{g_k\}$ converges to g almost everywhere on R. ■

Theorem 1-7: Almost uniform convergence $\Rightarrow$ Convergence in measure.

Proof: Given that a sequence $\{g_k\}$ on $(R, \Re, \mu)$ converges almost uniformly to g, let us choose $\varepsilon > 0$ and $\alpha > 0$. Then, $\exists$ $E_\varepsilon \in \Re$ with $\mu(E_\varepsilon) < \varepsilon$ s.t. $\{g_k\}$ converges uniformly to g on $R \setminus E_\varepsilon$. Therefore, if the positive integer $n(\alpha, \varepsilon)$ is made sufficiently large, then $\{t \in R : |g_k(t) - g(t)| \ge \alpha\} \subseteq E_\varepsilon$ implying convergence in measure. ■

Theorem 1-8 (Egoroff Theorem): For $P(R) < \infty$, Convergence almost everywhere $\Rightarrow$ Almost uniform convergence

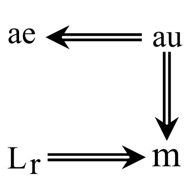
Proof: Let $\{g_k\}$ converge to g everywhere on $\tilde{R} \equiv R \setminus H$ where $P(H) = 0$. For $m, n \in \mathbf{N}$, let us construct a measurable set: $E_n^m \equiv \bigcup_{k=n}^{\infty} \left\{t \in \tilde{R} : |g_k(t) - g(t)| \ge \frac{1}{m}\right\}$ so that $E_{n+1}^m \subseteq E_n^m$ and, because of pointwise convergence of $\{g_k\}$ to g everywhere on $\tilde{R}$, we have $\bigcap_{n=1}^{\infty} E_n^m = \varnothing$. Since $P(R) < \infty$, we infer that $P\left(E_n^m\right) \to 0$ as $n \to \infty$.

For any given $\varepsilon > 0$, choose $\ell(m, \varepsilon) \in \mathbf{N}$ such that $P\left(E_\ell^m\right) < \varepsilon 2^{-m}$ and let $E_\varepsilon \equiv \bigcup_{m=1}^{\infty} E_\ell^m$ so that $P\left(E_\varepsilon\right) < \varepsilon$. Note that if $t \notin E_\varepsilon$, then $t \notin E_\ell^m$. Therefore, $|g_k(t) - g(t)| < \frac{1}{m}$ $\forall k \ge \ell$ implying that $\{g_k\}$ converges to g everywhere on $\tilde{R} \setminus E_\varepsilon$. ■

Corollary to Theorem 1-8: For $P(R) < \infty$, Convergence almost everywhere $\Rightarrow$ Convergence in measure P.

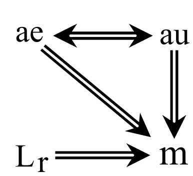
Proof: Proof follows by combining the results of Theorem 1-7 and Theorem 1-8. ■

Summary of Modes of Convergence



Finite or Infinite Measure

almost everywhere (ae)
almost uniformly (au)
in measure (m)
r[th] mean (L_r)



Finite Measure Only

Appendix A01: Introduction to Topology and Measure Theory

The notion of topology allows generalization of open sets and continuity of functions beyond metric spaces. Topology is a vast subject and therefore only the rudimentary concepts of point set topology are presented in this section. We also present rudimentary concepts of Measure Theory in this section. The objective is to bring in these concepts to those who have never been exposed to Measure Theory. Examples of easily readable books on these topics are Real Analysis by Royden (1989) and Elements of Integration and Lebesgue Measure by Bartle (1966). Also refer to Appendix D of Naylor and Sell (1982).

Topology and Topological Spaces

Definition A01-1: Let Ω be a nonempty set and let $\Im$ be a collection of subsets of Ω such that:

- $\varnothing \in \Im$ and $\Omega \in \Im$.
- If $S_k \in \Im$ for $k = 1,2,\cdots,n$, then $\bigcap_{k=1}^{n} S_k \in \Im$ finite intersection
- If $S_\alpha \in \Im$ for $\alpha \in I$ where I is the index set, then $\bigcup_{\alpha \in I} S_\alpha \in \Im$ arbitrary union

Then, $\Im$ is a topology of Ω; $\langle \Omega, \Im \rangle$ is a topological space; and each member of $\Im$ is said to be a $\Im$-open set in Ω. ♦

Definition A01-2: The usual topology $\langle \Re, \mathrm{U} \rangle$ is defined with $\Omega = \Re \equiv (-\infty, \infty)$ and $\Im = \mathrm{U}$ that contains all open intervals in $\Re$. A set $G \subseteq \Re$ is said to be U-open (i.e., open relative to the usual topology U) if either $G = \varnothing$ or, for $G \neq \varnothing$, $\forall p \in G$ $\exists$ an open interval $(a,b) \subset G$ such that $p \in (a,b)$. ♦

Definition A01-3: Let $\langle X, \Im \rangle$ be a topological space. Then the complement of every $\Im$-open set in X is said to be $\Im$-closed in X. That is, if $S \in \Im$, then $S^c \equiv X - S$ is $\Im$-closed in X. In other words, S is $\Im$-open in X if and only if $S^c \equiv X - S$ is $\Im$-closed in X. ♦

Definition A01-4: Let $\langle X, \Im \rangle$ be a topological space and let $p \in X$. Then, $B \subseteq X$ is called a $\Im$-neighborhood of $p \in X$ if $\exists$ a $\Im$-open set G such that $p \in G \subseteq B$.

Remark A01-1: Note that, in the topological sense, a $\Im$-neighborhood of a point $p \in X$ need not be a $\Im$-open set in X. However, a $\Im$-open set is a $\Im$-neighborhood of each of its points. ♦

Definition A01-5: Let $S \subset X$ where $\langle X, \Im \rangle$ is a topological space. Then, a point $p \in X$ is said to be a cluster point of S if every $\Im$-neighborhood of p contains at least one point of S other than p. In other words, p is a cluster point of S if and only if the following condition holds: B is a $\Im$-neighborhood of p implying that $(B - \{p\}) \cap S \neq \varnothing$.

<u>**Example A01-1**</u>: Consider the open interval $(0,1) \subset \Re$. In the usual topology $\langle \Re, \mathrm{U} \rangle$, both 0 and 1 are cluster points of $(0,1)$. Furthermore, every point of $(0,1)$ is a cluster point of $(0,1)$. ♦

Definition A01-6: Let $\langle X, \Im \rangle$ and $\langle Y, \vartheta \rangle$ be two topological spaces. Then a mapping $f: X \to Y$ is said to continuous (more precisely, $\Im - \vartheta$ continuous) if the inverse image $f^{-1}(\Phi)$ is $\Im$-open in X for every ϑ-open set Φ in Y. ♦

Definition A01-7: A $\Im-\vartheta$ continuous mapping $f: X \to Y$ is $\Im-\vartheta$-bicontinuous if $f[\Theta]$ is a $\vartheta-$open set in Y for every $\Im$-open set Θ in X. ♦

Definition A01-8: bijective and $\Im-\vartheta$-bicontinuous mapping $f: X \to Y$ is called a $\Im-\vartheta$-homeomorphism of X and Y. ♦

Remark A01-2: Two topological spaces are equivalent if they are homeomorphic. ♦

Definition A01-9: A topological space $\langle X, \Im \rangle$ is called a Hausdorff space (or a T_2-space) if, for every pair of distinct points x and y, i.e., $x, y \in X$ and $x \neq y$, $\exists$ $\Im$-neighborhoods B_x and B_y such that $B_x \cap B_y = \varnothing$. ♦

Example A01-2: The usual topology $\langle \Re, \mathsf{U} \rangle$ where the collection of all open subsets (defined in the usual sense) of $\Re = (-\infty, \infty)$ is Hausdorff. ♦

Example A01-3: Let $X = \{a,b,c\} \subset \Re$ and $\Im = \{\varnothing, \{a,b\}, \{c\}, X\}$. Clearly, the topological space is not Hausdorff because a and b are distinct points of X that do not have disjoint $\Im$-neighborhoods.

Definition A01-10: Let $\langle X, \Im \rangle$ be a topological space and $Y \subseteq X$. The $\Im$-relative topology of Y, denoted as $\Im_Y$, is defined as: $\Im_Y = \{G \cap Y : G \in \Im\}$. Then, $\langle Y, \Im_Y \rangle$ is called a subspace of $\langle X, \Im \rangle$.

♦

HW A01-1: Show that $\Im_Y$ is a topology of Y. ♦

Example A01-4: Let $X = \{a,b,c\} \subset \Re$ and $\Im = \{\varnothing, \{a,b\}, \{c\}, X\}$. Clearly, the topological space $\langle X, \Im \rangle$ is not Hausdorff because a and b are distinct points of X that do not have disjoint $\Im$-neighborhoods.

♦

Example A01-5: Let $Y = (0,1) \subset \Re$. Consider the relative topology $\langle Y, U_Y \rangle$ in which the interval $(\frac{1}{2},1)$ is open in $\langle Y, U_Y \rangle$. Although $[\frac{1}{2},1)$ is not closed in $\langle \Re, \mathsf{U} \rangle$ but $[\frac{1}{2},1)$ is closed in $\langle Y, U_Y \rangle$ because $[\frac{1}{2},1)$ is the complement of the U_Y-open set $(0,\frac{1}{2})$ in Y. The set $\left\{\frac{1}{k} : k \in \mathbf{N}\right\}$ is closed in the relative topology $\langle Y, U_Y \rangle$, $\langle \Re, \mathsf{U} \rangle$ because it has no cluster points in Y. However, $\left\{\frac{1}{\mathrm{k}} : \mathrm{k} \in \mathbf{N}\right\}$ is not closed in the usual topology $\langle \Re, \mathsf{U} \rangle$ because the cluster point 0 is not contained in $\left\{\frac{1}{k} : k \in \mathbf{N}\right\}$. ♦

Next, we present three important results:

Result A01-1: The topological spaces $\langle \Re, \mathsf{U} \rangle$ and $\langle (0,1), \mathsf{U}_{(0,1)} \rangle$ are homeomorphic. This result follows by constructing a bijective and bicontinuous function $f:(0,1) \to \Re$ such as $f(x) = \frac{2x-1}{x(x-1)}$.

♦

Result A01-2: If I_1 and I_2 are two U-open intervals in $\Re$, the spaces $\langle \mathrm{I}_1, \mathsf{U}_{\mathrm{I}1} \rangle$ and $\langle I_1, \mathsf{U}_{I2} \rangle$ are homeomorphic. ♦

Result A01-3: If I_1 and I_2 are two U-closed intervals in $\Re$, the spaces $\langle I_1, \mathsf{U}_{I1} \rangle$ and $\langle I_1, \mathsf{U}_{I2} \rangle$ are homeomorphic. ♦

Compactness in a Topological Space

Definition A01-11: A metric space $\langle S, d\rangle$ is sequentially compact if every sequence in S has a convergent subsequence. ♦

Example A01-6: The sequence $\{1, \frac{1}{2}, 3, \frac{1}{4}, 5, \frac{1}{6}, \cdots\}$ has a convergent subsequence $\left\{\frac{1}{2^k} : k \in \mathbf{N}\right\}$ in $\Re$. Note that the sequence $\{1, \frac{1}{2}, 3, \frac{1}{4}, 5, \frac{1}{6}, \cdots\}$ itself is not convergent in $\Re$ and that it contains many subsequences like $\{1, 3, 5, 7, \cdots\}$ which are not convergent. ♦

Example A01-7: The set $(0,1]$ is not compact in $\langle \Re, \mathsf{U}\rangle$ because the sequence $\left\{\frac{1}{k} : k \in \mathbf{N}\right\}$ does not have a subsequence with a limit point in $(0,1]$. ♦

Definition A01-12: Let $\langle S, \Im\rangle$ be a topological space and let $E \subseteq S$. Let $\Sigma = \{S_\alpha : \alpha \in I\}$ be a collection of subsets of S where I is an index set (which nonempty, and finite or countable or uncountable). Then, Σ is said to be a covering of E if $E \subseteq \bigcup_{\alpha \in I} S_\alpha$. If Σ_1 is a covering of E and Σ_2 is a covering of E such that $\Sigma_2 \subseteq \Sigma_1$, then Σ_2 is a subcovering of Σ_1. ♦

HW A01-2: Let $\Sigma_1 = \left\{(0, \frac{k}{k+1}) : k \in \mathbf{N}\right\}$. Verify that Σ_1 is a U-open covering of $(0,1)$ ♦

Example A01-8: If $\Sigma_2 = \left\{\left(0, \frac{4k+3}{4(k+1)}\right) : k \in \mathbf{N}\right\}$ is a subcovering of Σ_1, then $\Sigma_1 = \left\{(0, \frac{k}{k+1}) : k \in \mathbf{N}\right\}$ ♦

Definition A01-13: Let $\langle S, \Im\rangle$ be a topological space. A covering Σ of $E \subseteq S$ is said to be a $\Im$-open covering of E if every member of Σ is a $\Im$-open set. A covering Σ of a set E is said to be finite if card(Σ) is finite. ♦

Definition A01-14: A topological space $\langle S, \Im\rangle$ is said to be compact if every $\Im$-open covering of S has a finite subcovering. ♦

Theorem A01-1: The topological space $\langle \Re, \mathsf{U}\rangle$ is not compact. Therefore, no open interval on $\Re$ is compact in its relativized U-topology.

Proof of Theorem A01-1: Let $\Sigma = \{(-k, k) : k \in \mathbf{N}\}$. Then, Σ is a U-open covering of $\Re$ because each member of Σ is an open interval in $\Re$, and $\Re \subset \bigcup_{k \in \mathbf{N}} (-k, k)$. If $x \in \Re$, then $\exists\, n_x \in \mathbf{N}$ such that $n_x > |x| \Rightarrow x \in (-n_x, n_x)$. So, $x \in \bigcup_{n \in \mathbf{N}} (-n, n)$. Now, let $(-n_1, n_1), (-n_2, n_2), \cdots, (-n_k, n_k)$ be any finite collection of members of Σ and let $n^* = \max(n_1, n_1, \cdots, n_k)$. Then, $n^* \notin \bigcup_{k \in \mathbf{N}} (-n_k, n_k)$. Therefore, there is no finite collection of members of Σ which is a covering of $\Re$. So, $\langle \Re, \mathsf{U}\rangle$ is not compact. The second assertion follows from homeomorphism between $\langle \Re, \mathsf{U}\rangle$ and $\langle \mathbf{I}, \mathsf{U}\rangle$ where $\mathbf{I}$ is an interval in $\Re$. ♦

Definition A01-15: A mapping $f : X \to \Re$ is bounded if the range of $f[X]$ is a bounded subset of $\Re$. ♦

Next we present the following important results without proof.

Result A01-4: If $\langle X, \Im_X \rangle$ is a compact subspace of a Hausdorff space $\langle \Omega, \Im \rangle$, then X is $\Im$-closed. ♦

Result A01-5: If $\langle \Omega, \Im \rangle$ is compact and X is a $\Im$-closed subset of X , then $\langle X, \Im_X \rangle$ is compact. ♦

Result A01-6: (Heine-Borel Theorem) For $X \subset \Re$, $\langle X, \mathsf{U}_X \rangle$ is compact iff X is bounded and U-closed. ♦

Result A01-7: A continuous image of a compact space is compact. That is, for two topological spaces $\langle X, \Im \rangle$ and $\langle Y, \vartheta \rangle$, if $f : X \to Y$ is $\Im - \vartheta$-continuous, compactness of $\langle X, \Im \rangle$ implies compactness of $\langle Y, \vartheta \rangle$. ♦

Result A01-8: Let $\langle X, \Im \rangle$ be a compact space and let $\langle Y, \vartheta \rangle$ be a Hausdorff space. If $f : X \to Y$ is $\Im - \vartheta$-continuous and surjective, then f is a homeomorphism. ♦

Result A01-9: Let $\langle X, \Im \rangle$ be a compact space. If $f : X \to \Re$ is $\Im - \mathsf{U}$-continuous, then $f(\cdot)$ is bounded. ♦

Result A01-10: (Bolzano-Weierstrass Theorem): Every bounded infinite subset of $\Re$ has at least one U-cluster point. ♦

Result A01-11: Let $X \subseteq \Re$ be bounded and U-closed. If $f : X \to \Re$ is $\mathsf{U}_X - \mathsf{U}$-continuous, then f is bounded.

♦

Total Boundedness and Approximation

Definition A01-16: Let E be a set in a metric space $\langle S, d \rangle$. Given $\varepsilon > 0$, $E_\varepsilon \subset E$ is an $\varepsilon -$ net of E if :

(i) E_ε is a finite set; and (ii) $\forall x \in E, \exists y \in E_\varepsilon$ such that $d(x, y) < \varepsilon.$ ♦

Definition A01-17: A set E in a metric space $\langle S, d \rangle$ is totally bounded if: $\forall \varepsilon > 0, \exists$ an $\varepsilon -$ net in E.

♦

Remark A01-3: Total boundedness implies boundedness. The converse is true for all finite-dimensional spaces but, in general, it is not true for infinite-dimensional spaces. ♦

Remark A01-3: Every finite set in a metric space is bounded and hence it is totally bounded. ♦

Example A01-8: Consider the closed ball $\widetilde{B}_1(\underline{0}) = \{x \in \ell_2 : \|x\|_{\ell_2} \leq 1\}$ where the distance function is defined as:

$$d(x, y) = \|x - y\|_{\ell_2} \equiv \sqrt{\sum_{k=1}^{\infty} |x_k - y_k|^2} \ \forall x, y \in \ell_2$$

The set $\widetilde{B}_1(\underline{0})$ is bounded because $d(x, y) \leq 2 \ \forall x, y$ but $\widetilde{B}_1(\underline{0})$ is not totally bounded as seen below. Let us construct $E = \{e^k : k \in \mathbf{N}\}$ where e^k is the sequence of all 0's except '1' as the k^{th} element of the sequence. Clearly, $d(e^k, e^\ell) = \sqrt{2}\delta_\ell^k$. If an $\varepsilon -$ net $E_{1/2}$ exists for $\varepsilon = \frac{1}{2}$, then $E_{1/2}$ must be a finite subset in E. But since the closed balls $\widetilde{B}_{1/2}(e^k)$ and $\widetilde{B}_{1/2}(e^\ell)$ are disjoint for all $k \neq \ell$, $E_{1/2}$ must contains a point within a distance $\frac{1}{2}$ of each e^k. Since there are countably many e^k's, $E_{1/2}$ cannot be a finite set. However, notice that this violation of finite cardinality would not have occurred in a finite-dimensional space.

Next we present the following important results without proof.

Result A01-12: Let $E \subseteq X$ in a metric space $\langle X, d \rangle$. If E has an $\varepsilon-$net for some $\varepsilon > 0$, then E is bounded. ♦

Result A01-13: Let $\langle X, d \rangle$ be a totally bounded metric space. Then, X is separable. ♦

Result A01-14: A bounded set $E \subseteq \ell_2$ is totally bounded iff $\forall \varepsilon > 0 \, \exists \, n(\varepsilon) \in \mathbb{N}$ such that

$$\sum_{k=n}^{\infty} |x_k|^2 < \varepsilon \;\; \forall x \in E$$ ♦

Result A01-15: Let $E \subseteq X$ in a metric space $\langle X, d \rangle$. Then, the following statements are equivalent:

(i) The closure $\overline{E}$ is sequentially compact.

(ii) Every sequence in E has a subsequence that converges in X. ♦

Result A01-16: Every sequentially compact set in a metric space is closed. ♦

Result A01-17: Every sequentially compact metric space is complete. ♦

Result A01-18: Every sequentially compact metric space is totally bounded. ♦

Result A01-19: A metric space is sequentially compact iff it is totally bounded and complete. ♦

CONCEPT OF MEASURE AND MEASURABLE SPACES

Intuitively, Lebesgue measure is the length of an interval or of an at most countable union of intervals on the real line $\Re$. This concept can be extended to $\Re^2$ and $\Re^3$ as areas and volumes, and also to other finite-dimensional spaces. The concept of axiomatic probability theory, introduced by Kolmogorov, is based on the principle of Lebesgue measure. In general, measure is a set function, i.e., an assignment of a number m(A) to each set in a certain class.

Consider the open interval (a, b) whose measure is the length $(b-a)$. Similarly, the length of a countable union of disjoint open intervals can be obtained by summing the lengths of these intervals. Defining sets only in terms of disjoint open intervals is often restrictive. Therefore, we would like to generalize the concept of measure. At this stage, solely for simplicity, we would restrict the treatment of measure to finite intervals and bounded subsets of $\Re$.

- For a finite interval $\mathbf{I}$ (open, closed or semi-open), the measure is equal to the length of the interval, i.e., $m(\mathbf{I}) = \ell(\mathbf{I})$.
- If $\{\mathbf{I_k}\}$ is an at most countable (i.e., finite or countably infinite) sequence of disjoint intervals, then $m\left(\bigcup_{i=1}^{\infty} \mathbf{I}_i\right) = \sum_{i=1}^{\infty} \ell(\mathbf{I}_i)$.
- The measure is translation-invariant. That is, if E is a set for which the measure m is defined and if, for any given $y \in \Re$, $E \oplus y$ is the set $\{x + y : x \in E\}$ obtained by replacing each point x in E by $x + y$, then $m(E \oplus y) = m(E)$.

<u>Outer Measure</u>: For each set $E \subseteq \Re$, consider the countable collection $S = \{\mathbf{I}_k\}$ of open intervals in $\Re$ that cover E, i.e., $E \subseteq \bigcup_{i=1}^{\infty} \mathbf{I}_i$. The outer measure of E is then defined as:

$$\overline{m}(E) = \inf_{all\,S} \sum_{j=1}^{\infty} \ell(\mathbf{I}_j)$$

Clearly, if $A \subseteq B$, then $\overline{m}(A) \leq \overline{m}(B)$. Also, $\overline{m}(\varnothing) = 0$ and each set consisting of a single point has a zero outer measure. The following results are presented from Royden (1989) without proof:

- <u>Result 1</u>: The outer measure of an interval (open or closed or semi-open) is its length.

- Result 2: If $\{A_k\}$ is a sequence of subsets of $\Re$, then $\overline{m}\left(\bigcup_{k=1}^{\infty} A_k\right) \le \sum_{k=1}^{\infty} \overline{m}(A_k)$.
- Result 3: If A is a countable subset (dense or not) of $\Re$, then $\overline{m}(A)$ = 0.
- Result 4: If A is an uncountable set, then $\overline{m}(A) \ge 0$. Usually the measure of an uncountable set is greater than 0. However, there are uncountable sets such as the Cantor set whose measure is 0.

Definition A01-18: A set $E \subseteq \Re$ is said to be measurable if, for every $A \subseteq \Re$, the following condition holds: $\overline{m}(A) = \overline{m}(A \cap E) + \overline{m}(A \cap E^c)$. ♦

The fact that $A = (A \cap E) \cup (A \cap E^c)$ implies, by virtue of Result 2, that $\overline{m}(A) \le \overline{m}(A \cap E) + \overline{m}(A \cap E^c)$. Therefore, E is measurable whenever $\overline{m}(A) \ge \overline{m}(A \cap E) + \overline{m}(A \cap E^c)$. Furthermore, because of symmetry, E^c is measurable if and only if E is measurable.

Inner Measure: Assuming that E is a bounded set, we choose an interval $\mathbf{I}^*$ such that $E \subset \mathbf{I}^*$. We define the inner measure $\underline{m}$ as: $\underline{m}(E) = \ell(\mathbf{I}^*) - \overline{m}(\mathbf{I}^* - E)$. The following results are presented without proof:

- Result 5: The inner measure $\underline{m}(E)$ is invariant for every $\mathbf{I}^*$ containing E.
- Result 6: For every bounded set E, then $\underline{m}(E) \le \overline{m}(E) < \infty$.
- Result 7: If E is a finite interval, then $\underline{m}(E) = \overline{m}(E) = \ell(E)$.
- Result 8: A bounded set $E \subseteq \Re$ is measurable if $\underline{m}(E) = \overline{m}(E)$. In that case, we denote the measure as: $m(E) = \underline{m}(E) = \overline{m}(E)$.

Demonstration of the Existence of a Nonmeasurable Set:

The usually encountered sets are measurable and, for engineering applications, we may not have to deal with any nonmeasurable sets. Indeed, it is not easy to find a nonmeasurable set. However, from the conceptual point of view, it is important to establish the existence of a non-measurable set. An example of a nonmeasurable set [Royden (1989), pp. 64-65] is given below.

Let $x, y \in [0,1)$. Define the sum Modulo 1 as: $x \oplus y = \begin{cases} x+y & \text{if } (x+y) < 1 \\ x+y-1 & \text{if } (x+y) \ge 1 \end{cases}$

where the operator $\oplus$ can be interpreted as follows: If $\theta = x \oplus y$, then the angle $2\pi\theta$ in radians is the sum modulo addition of two angles $2\pi x$ and $2\pi y$. The sum modulo operator $\oplus$ can also be translated, i.e., $E \oplus y = \{z : z = x \oplus y \text{ for } x \in E\} \ \forall y \in E$. Furthermore, the operator $\oplus$ is commutative and associative. We establish the following lemma before citing an example of a nonmeasurable set.

Lemma: Let $E \subseteq [0,1)$ be a measurable set. Then, the set $E \oplus y$ is measurable and m($E \oplus y$) = m(E) $\forall y \in E$.

Proof: Let $E_1 = E \cap [0, 1-y)$ and $E_2 = E \cap [1-y, 1)$ for some $y \in [0,1)$.. Therefore, $E_1 \cap E_2 = \varnothing$ and $E_1 \cup E_2 = E$ which imply that $m(E_1) + m(E_2) = m(E)$. Since the set $E_1 + y$ is measurable and $E_1 \oplus y = E_1 + y$, we argue that $E_1 \oplus y$ is also measurable and m($E_1 \oplus y$) = m($E_1 + y$) = m(E_1). Similarly, $E_2 \oplus y = E_2 + y - 1$ implies that $E_2 \oplus y$ is measurable and m($E_2 \oplus y$) = m(E_2). But $(E_1 \oplus y) \cup (E_2 \oplus y) = E \oplus y$ and $(E_1 \oplus y) \cap (E_2 \oplus y) = \varnothing$. Hence,

$m((E \oplus y)) = m(E_1 \oplus y) + m(E_2 \oplus y) = m(E_1) + m(E_2) = m(E)$.

Now we proceed to construct a nonmeasurable set. Let $x, y \in [0,1)$ such that $(x-y)$ is rational. This is an equivalence relation $x \sim y$ because: (i) $x \sim x$ since 0 is rational; (ii) $x \sim y \equiv y \sim x$ since if r is rational, so is $-r$; and (iii) ($x \sim y$ and $y \sim z$) $\Leftrightarrow x \sim z$ because if r and $\tilde{r}\, r'$ are rational, so is $r + \tilde{r}$. Next, we partition the set $[0,1)$ into equivalence classes such that any two elements of a given equivalence class differ by a rational, and any two numbers belonging to different equivalence classes differ by an irrational.

Let us construct the set P that contains exactly one number from each equivalence class and assume that P is a measurable set. Let $r_i,\ i = 0,1,2,\cdots$, be an enumeration of the rational numbers in $[0,1)$ with $r_0 = 0$. Let us define $P_i = P \oplus r_i \Rightarrow P_0 = P$ and let $x \in P_i \cap P_j$ for $i \neq j$. Therefore, $x = q_i \oplus r_i = q_j \oplus r_j$ with $q_i, q_j \in P$. Then, $q_i - q_j = r_i - r_j$ is a rational and hence $q_i \sim q_j$. Since P has exactly one element from each equivalence class, $q_i = q_j$. That means, for $i \neq j$, $P_i \cap P_j = \varnothing$. Therefore, the collection of sets $\{P_i\}$ is pairwise disjoint. On the other hand, each $x \in [0,1)$ belongs to one and only one of the equivalence classes and therefore must be equivalent to an element of P. But, if x differs from an element of P by a rational r_i, then $x \in P_i$ for some $i \in \{1,2,3,\cdots\}$. Therefore, $\bigcup_{i=1}^{\infty} P_i = [0,1)$. Further, since P_i is a translation modulo 1 of P, we conclude by the lemma, that each P_i is measurable and has the same measure as P. If it is so, $m([0,1)) = \sum_{i=0}^{\infty} m(P_i) = \sum_{i=0}^{\infty} m(P)$ which implies that $m([0,1))$ is either 0 or ∞ depending on whether $m(P)$ is zero or non-zero. But we know that $m([0,1)) = 1$. This is a contradiction. So P is a nonmeasurable set. ♦

Measurable Functions and Convergence almost everywhere

Consider a sequence of functions $\{f_k\}$ defined on a set E. If $\{f_k\}$ converges to a function f at every point on E except possibly on a set $A \subset E$ where $m(A) = 0$, then $\{f_k\}$ is said to converge to f almost everywhere, abbreviated as *a.e.*, on E. The a.e. convergence is conceptually similar to the almost sure (a.s.) convergence of a random sequence.

Definition A01-19: A collection Ψ of subsets of a non-empty set Ω (which may be finite or countably infinite or uncountable) is said to be an algebra in Ω if Ψ satisfies the following properties:

(i) $\Omega \in \Psi$.

(ii) If $E \in \Psi$, then $E^c \in \Psi$ where $E^c \equiv \Omega - E$.

(iii) $E = \bigcup_{i=1}^{n} E_i \in \Psi$ where $E_i \in \Psi\ \forall$i, then $E \in \Psi$. [This property is as closure under finite union.]

If condition (iii) is relaxed to countable union, i.e., if $E = \bigcup_{i=1}^{\infty} E_i \in \Psi$, then Ψ is called a σ-algebra in Ω.

The duple (Ω, Ψ) is called a *measurable space* where the members of Ψ are called measurable sets in Ω. When there is no confusion, we say Ω is a measurable space instead of (Ω, Ψ). ♦

Remark A01-20: The last two conditions imply that any finite (countable) intersections of events is also an event for an algebra (σ-algebra). ♦

Remark A01-21: In the terminology of probability theory, the non-empty set Ω is called the sample space which is the set of all possible outcomes (of a random experiment) or sample points, and $E \in \Psi$ is called an event. In general, the σ-algebra Ψ is called the event space which is the collection of all possible events. It should be obvious from the above three conditions that any arbitrary subset of Ω may not be qualified as an event. However, the sample space Ω (which is the sure event) and its complement in Ω, namely the empty set $\varnothing$, (which is called the impossible event) are always qualified as events. Every event space must contain these two events. Therefore, for a given sample space, the event space may not be unique. So, the smallest event space which can be obtained as the intersection of all possible event spaces is $\{\varnothing, \Omega\}$. ♦

Remark A01-22: If Ω is a finite set, then there can be only finitely many event spaces, each of which must also be a finite set. In other words, there can be only finitely many different algebras if there are only finitely many elements in Ω. The largest possible event space is the power set 2^{Ω}. However, if the cardinality of Ω is 1, i.e., if there is exactly one experimental outcome, then the only possible event space is $\{\varnothing, \Omega\}$. ♦

Remark A01-23: If Ω is an infinite set, then Ψ can be finite or infinite. This follows from the facts that the smallest Ψ is always finite and the largest Ψ is the power set 2^{Ω} which is infinite if Ω is infinite. Note that, for an infinite Ω, countable or uncountable, it is possible to construct an uncountable Ψ but there does not exist a countably infinite Ψ. ♦

Remark A01-24: It follows from De Morgan's theorem and the last two conditions of σ-algebra that any countable intersection of events is also an event, i.e., if $E = \bigcap_{i=1}^{\infty} E_i \in \Psi$ if $E_i \in \Psi$. ♦

Remark A01-25: In the context of probability theory, each event (i.e., element of the event space) is a measurable set. ♦

Definition A01-20: A nonnegative finitely additive set function μ defined on $\mathfrak{E}$ is called finite iff μ(Ω) is finite. This implies that μ(E) is finite for every $E \in \mathfrak{E}$. Furthermore, μ defined on $\mathfrak{E}$ is called σ-finite iff there exist a sequence $\{E_i\}$ with $E_i \in \mathfrak{E}$ such that $\Omega = \bigcup_{i=1}^{\infty} E_i$ and $\mu(E_i) < \infty \; \forall i$. (Note that Lebesgue measure is σ-finite but not finite.). ♦

Appendix A02: Important Theorems for Exchange of Limits, Summation, and Integrals

In many engineering problems, we exchange the orders of limits, infinite sums, and integrals. One should be cautious about these exchanges because such operations may not be always valid and may cause errors under certain circumstances. (Note that there is no problem in exchanging finite summations with limits, infinite summations, and integrals.) Before stating relevant theorems to support these notions, we cite a few examples to demonstrate that such problems do exist in engineering analysis.

Case 1: In general, $\lim_{n\to\infty}\lim_{m\to\infty} f_{m,n} \neq \lim_{m\to\infty}\lim_{n\to\infty} f_{m,n}$

We cite an example. Let $f_{m,n} = \dfrac{m}{m+n}$. Then, for any fixed $n \in \mathbf{N}$, we have

$$\lim_{m\to\infty} f_{m,n} = \lim_{m\to\infty} \frac{m}{m+n} = 1 \;\Rightarrow\; \lim_{n\to\infty}\lim_{m\to\infty} f_{m,n} = 1.$$

However, for any fixed $m \in \mathbf{N}$, we have:

$$\lim_{n\to\infty} f_{m,n} = \lim_{n\to\infty} \frac{m}{m+n} = 0 \;\Rightarrow\; \lim_{n\to\infty}\lim_{m\to\infty} f_{m,n} = 0.$$ ♦

Case 2: In general, $f_n \to f \Rightarrow \dot{f}_n \to \dot{f}$ where $\dot{f}(t)$ indicates $\frac{d}{dt}\left(f(t)\right)$.

We cite an example. Let $f_n(t) = \dfrac{\sin(nt)}{\sqrt{n}}$ for $t \in \mathbf{R} \equiv (-\infty,\infty)$ and $n \in \mathbf{N}$. Then,

$$f(t) = \lim_{n\to\infty} f_n(t) = \lim_{n\to\infty} \frac{\sin(nt)}{\sqrt{n}} = 0 \;\Rightarrow\; \dot{f}(t) \equiv \tfrac{d}{dt}\left(f(t)\right) = 0$$

But $f_n'(t) \equiv \frac{d}{dt}\left(f_n(t)\right) = \frac{d}{dt}\left(\dfrac{\sin(nt)}{\sqrt{n}}\right) = \sqrt{n}\cos(nt) \Rightarrow \lim_{n\to\infty} \dot{f}_n(t) = \lim_{n\to\infty} \sqrt{n}\cos(nt)$ does not exist in $\mathbf{R}$. ♦

Case 3: In general, $\lim\int \neq \int\lim$

We cite an example. Let $f_n(t) = n^2 t\left(1-t^2\right)^n$ for $t \in [0,1]$ and $n \in \mathbf{N}$. Then,

$$f(t) \equiv \lim_{n\to\infty} f_n(t) = \lim_{n\to\infty} n^2 t\left(1-t^2\right)^n = 0 \quad \forall t \in [0,1]$$

$$\text{and } \int_0^1 dt\, f(t) = \int_0^1 dt \lim_{n\to\infty} n^2 t(1-t^2)^n = 0$$

On the other hand, $\int_0^1 dt\, f_n(t) = \int_0^1 dt\, n^2 t\left(1-t^2\right)^n = \frac{n^2}{2(n+1)} \;\Rightarrow\; \lim_{n\to\infty}\int_0^1 dt\, f_n(t) = \lim_{n\to\infty} \frac{n^2}{2(n+1)}$ does not exist in $\mathbf{R}$.

We cite another example where both limits exist but they are unequal. Let $f_n(t) = nt\left(1-t^2\right)^n$ for $t \in [0,1]$, $n \in \mathbf{N}$.

Then, $f(t) \equiv \lim_{n\to\infty} f_n(t) = \lim_{n\to\infty} nt\left(1-t^2\right)^n = 0 \quad \forall t \in [0,1] \Rightarrow \lim_{n\to\infty}\int_0^1 dt\, f_n(t) = 0$.

On the other hand, $\int_0^1 dt\, f_n(t) = \int_0^1 dt\, nt\left(1-t^2\right)^n = \frac{n}{2(n+1)} \;\Rightarrow\; \lim_{n\to\infty}\int_0^1 \mathrm{dt}\, \mathrm{f_n(t)} = \lim_{n\to\infty} \frac{n}{2(n+1)} = \frac{1}{2}$.

Both limits exist but they are unequal. ♦

Case 4: In general, $\int_{T_x} dx \int_{T_y} dy\, f(x,y) \neq \int_{T_y} dy \int_{T_x} dx\, f(x,y)$

We cite an example. Let $f(x,y) = \dfrac{x-y}{(x+y)^3}$ for $x \in [0,1]$ and $y \in [0,1]$. Then,

$$\int_0^1 dx \int_0^1 dy \left(\frac{x-y}{(x+y)^3} \right) = \int_0^1 dx \left[\frac{1}{x+y} - \frac{x}{(x+y)^2} \right]_{y=0}^1 = \int_0^1 \frac{dx}{(1+x)^2} = \frac{-1}{(1+x)}\bigg|_{x=0}^1 = \frac{1}{2} \text{ and}$$

$$\int_0^1 dy \int_0^1 dx \left(\frac{x-y}{(x+y)^3} \right) = \int_0^1 dy \left[\frac{y}{(x+y)^2} - \frac{1}{x+y} \right]_{x=0}^1 = -\int_0^1 \frac{dy}{(1+x)^2} = \frac{1}{(1+y)}\bigg|_{y=0}^1 = -\frac{1}{2}$$

The reasons for inequality of these two integrals are that (i) the function $f \notin L_1(T_x \times T_y)$, i.e., f is not absolutely integrable on $[0,1] \times [0,1]$; and (ii) f changes sign on its range, i.e., f becomes negative on its range. ♦

We now state several important theorems, without proof, that are important for determining when exchange of limits, infinite summation, and integrals are permissible.

Theorems Stated in Non-Measure-Theoretic Terms

Theorem A02-1: Let $\alpha(\bullet)$ be a monotonically increasing function on $\mathbf{T} \subseteq \mathbf{R}$, i.e., $\alpha(x) \geq \alpha(y)\ \forall x > y$. Let $\{\varphi_k\}$ be a sequence of functions that Riemann-Stieltjes integrable on w.r.t. α. If $\varphi_k \to \varphi$ uniformly on $\mathbf{T}$, then

- φ is Riemann-Stieltjes integrable on $\mathbf{T}$ w.r.t. α.
- $\lim_{k\to\infty} \int_T d\alpha(t)\varphi_k(t) = \int_T d\alpha(t)\varphi(t)$ (i.e., $\lim\int = \int\lim$) ♦

Corollary to Theorem A02-1: If the series $s_k(t) \equiv \sum_{n=1}^{k} \varphi_n(t)$ converges uniformly to $s(t)$ on $\mathbf{T}$, then

- $\lim_{k\to\infty} \sum_{n=1}^{k} \int_{\mathbf{T}} d\alpha(t)\varphi_n(t) = \int_{\mathbf{T}} d\alpha(t)s(t)$ (i.e., $\sum_1^\infty \int = \int \sum_1^\infty$) ♦

Theorem A02-2: Let $\varphi_k \to \varphi$ uniformly on $\mathbf{T} \subseteq \mathbf{R}$, and let τ be a limit point on $\mathbf{T} \subseteq \mathbf{R}$. If $\lim_{t\to\tau} \varphi_k(t) = \nu_k\ \forall k \in N$, then the sequence $\{\nu_k\}$ converges and $\lim_{t\to\tau} \varphi(t) = \lim_{t\to\tau} \nu_k$. In other words, $\lim_{t\to\tau} \lim_{k\to\infty} \varphi_k(t) = \lim_{k\to\infty} \lim_{t\to\tau} \varphi_k(t)$. ♦

Theorem A02-3: Let $\{\varphi_k\}$ be a sequence of functions that are differentiable on $\mathbf{T} \subseteq \mathbf{R}$, and let $\{\varphi_k(\tau)\}$ converge to $\varphi(\tau)$ for some $\tau \in \mathbf{T}$. If the derivative $\dot{\varphi}_k(\tau) \equiv \frac{d\varphi_k(t)}{dt}\Big|_{t=\tau}$ converges uniformly on $\mathbf{T}$, then

- $\{\varphi_k\}$ converges to φ uniformly on $\mathbf{T}$.
- $\lim_{k\to\infty} \dot{\varphi}_k(t) = \dot{\varphi}(t)$ on $\mathbf{T}$ (i.e., $\lim \frac{d}{dt} = \frac{d}{dt} \lim$) ♦

Theorem A02-4 (Fubini's Theorem): Let $\varphi : \mathbf{R}^{\ell} \times \mathbf{R}^{m} \to \mathbf{R}.$. Let $\varphi^{y}(x) = \varphi(x,y)$ for any fixed $y \in \mathbf{R}^{m}$, and $\varphi_{x}(y) = \varphi(x,y)$ for any fixed $x \in \mathbf{R}^{\ell}$. Then, the following conditions (i) and (ii) hold:

(i) If $\left(\int_{\mathbf{R}^m} d\mu(\mathrm{x}) \int_{\mathbf{R}^{\ell}} \mathrm{d}\nu(\mathrm{y}) \left|\varphi_x(y)\right| < \infty \right)$ or if $\left(\int_{\mathbf{R}^m} dy \int_{\mathbf{R}^{\ell}} d\mathrm{x} \left|\varphi(x,y)\right| < \infty \right)$, then

(a) $\varphi_x \in L_1(\mathbf{R}^m)$ and $\zeta(x) \equiv \int_{\mathbf{R}^m} dy\, \varphi_x(y) \in L_1(\mathbf{R}^{\ell})$.

(b) $\varphi^y \in L_1(\mathbf{R}^{\ell})$ and $\psi(y) \equiv \int_{\mathbf{R}^{\ell}} dx\, \varphi^y(x) \in L_1(\mathbf{R}^m)$.

(c) $\int_{\mathbf{R}^{\ell}} dx \int_{\mathbf{R}^m} \mathrm{dy} \left|\varphi(x,y)\right| = \int_{\mathbf{R}^m} dy \int_{\mathbf{R}^{\ell}} \mathrm{dx} \left|\varphi(x,y)\right| = \int_{\mathbf{R}^{\ell} \times \mathbf{R}^m} dx\, dy \left|\varphi(x,y)\right|$.

(ii) If $\varphi(x,y) \geq 0$ almost everywhere on $X \times Y$, then

$$\int_{\mathbf{R}^{\ell}} dx \int_{\mathbf{R}^m} \mathrm{dy} \left|\varphi(x,y)\right| = \int_{\mathbf{R}^m} dy \int_{\mathbf{R}^{\ell}} \mathrm{dx} \left|\varphi(x,y)\right| = \int_{\mathbf{R}^{\ell} \times \mathbf{R}^m} dx\, dy \left|\varphi(x,y)\right|.$$ ♦

Theorems Stated in Measure-Theoretic Terms

Definition: A function $\varphi : X \to [0,\infty]$ is defined to be integrable over a μ-measurable set E if $\int_E |\varphi| d\mu < \infty$.

Theorem A02-5 (Lebesgue-Monotone Convergence Theorem): Let $\{\varphi_k : X \to [0,\infty]\}$ be a sequence of monotonically increasing non-negative Lebesgue-measurable functions on a measure space (X,Σ,μ) such that $\lim_{k\to\infty} \varphi_k = \varphi$ almost everywhere on X. Then,

- φ is a Lebesgue-measurable function.
- $\lim_{k\to\infty} \int_X d\mu\, \varphi_k = \int_X d\mu\, \varphi$ (i.e., $\lim\int = \int\lim$) ♦

Corollary to Theorem A02-5: Let $\{\varphi_k : X \to [0,\infty]\}$ almost everywhere on X be a sequence of Lebesgue-measurable functions on a measure space (X,Σ,μ), and let $s_k \equiv \sum_{n=1}^{k} \varphi_n$ such that $\lim_{k\to\infty} s_k = s$ almost everywhere on X. Then,

- $\lim_{k\to\infty} \sum_{n=1}^{k} \int_X d\mu\, \varphi_n = \int_X d\mu\, s$ (i.e., $\sum_1^{\infty}\int = \int\sum_1^{\infty}$) ♦

Theorem A02-6 (Fatou's Lemma): Let $\{\varphi_k : X \to [0,\infty]\}$ be a sequence of Lebesgue-measurable functions on a measure space (X,Σ,μ) such that there exists an integrable function $\varphi : X \to [0,\infty]$ having the property that $\varphi_k \leq \varphi$ almost everywhere on X. Then,

- $\int_X d\mu \liminf \varphi_k \leq \liminf \int_X d\mu\, \varphi_k$ ♦

Theorem A02-7 : Let $\varphi : X \to [0,\infty]$ almost everywhere on X be a Lebesgue-measurable function on a measure space (X,Σ,μ), and let $\psi(E) \equiv \int_E d\mu\, \varphi \quad \forall E \in \Sigma$. Then,

- ψ is a measure on Σ.

- $\int_X d\psi\, f = \int_X d\mu\, f\varphi$ for every Lebesgue-measurable function f on a measure space (X, Σ, μ). ♦

Theorem A02-8 : Let φ be a Lebesgue-measurable function on a measure space (X, Σ, μ). If $\varphi \in L_1(\mu)$, i.e., φ is absolute-integrable w.r.t. the measure μ. Then, $\left|\int_E d\mu\, \varphi\right| \le \int_E d\mu\, |\varphi| \quad \forall E \in \Sigma.$ ♦

Theorem A02-9 (Lebesgue Dominated Convergence Theorem): Let $\{\varphi_k\}$ be a sequence of Lebesgue-measurable functions on a measure space (X, Σ, μ) such that there exists an integrable function $\lim_{k\to\infty} \varphi_k = \varphi$ a.e. on X. If $\exists\ \psi \in L_1(\mu)$ such that $|\varphi_k| \le \psi \quad \forall k \in \mathbf{N}$ almost everywhere on X, then

- $\varphi \in L_1(\mu)$
- $\lim_{k\to\infty} \int_X d\mu |\varphi_k - \varphi| = 0$
- $\lim_{k\to\infty} \int_X d\mu\, \varphi_k = \int_X d\mu\, \varphi$ (i.e., $\lim\int = \int\lim$) ♦

Theorem A02-10 (Fubini's Theorem): Let (X, Σ, μ) and (Y, T, ν) be $\sigma-$finite measure spaces. Let φ be a $(\Sigma \times \mathrm{T})$-measurable function on $X \times Y$ with the product measure $\pi \equiv \mu \times \nu$. Let $\varphi^y(x) \equiv \varphi(x, y)$ for any fixed $y \in Y$, and $\varphi_x(y) \equiv \varphi(x, y)$ for any fixed $x \in X$. Then, the following conditions (i) and (ii) hold:

(i) If $\left(\int_X d\mu(\mathrm{x}) \int_Y \mathrm{d}\nu(\mathrm{y}) |\varphi_x(y)| < \infty\right)$ or if $\left(\int_Y d\nu(y) \int_X d\mu(\mathrm{x}) |\varphi(x, y)| < \infty\right)$, then

(a) $\varphi_x \in L_1(\nu)$ and $\zeta(x) \equiv \int_Y d\nu(y) \varphi_x(y) \in L_1(\mu)$.

(b) $\varphi^y \in L_1(\mu)$ and $\psi(y) \equiv \int_X d\mu(x) \varphi^y(x) \in L_1(\nu)$.

(c) $\int_X d\mu(\mathrm{x}) \int_Y \mathrm{d}\nu(\mathrm{y}) |\varphi(x, y)| = \int_Y d\nu(y) \int_X d\mu(\mathrm{x}) |\varphi(x, y)| = \int_{X\times Y} d\pi(x, y)\, |\varphi(x, y)|$.

(ii) If $\varphi(x, y) \ge 0$ almost everywhere on $X \times Y$, then

$\int_X d\mu(\mathrm{x}) \int_Y \mathrm{d}\nu(\mathrm{y}) |\varphi(x, y)| = \int_Y d\nu(y) \int_X d\mu(\mathrm{x}) |\varphi(x, y)| = \int_{X\times Y} d\pi(x, y)\, |\varphi(x, y)|$. ♦

Appendix B: Conditional Expectation (Note Set # 1)

Let $(\Omega, \mathcal{E}, P)$ be a probability space and let $X : \Omega \to \mathbb{R}^n$ be a measurable (vector-valued) function such that $E[\|X_k\|] < \infty \ \forall k$. Let $\mathcal{H} \subset \mathcal{E}$ be a σ-algebra and let us denote the conditional expectation of X given $\mathcal{H}$ as $E[X|\mathcal{H}]$.

<u>Definition B-1</u>: The function $E[X|\mathcal{H}] \to \mathbb{R}^n$ is defined as:

$$E[X|\mathcal{H}] \text{ is } \mathcal{H}\text{-}\mathbb{R}^n\text{-measurable} \qquad (1)$$

and $$\int_H E[X|\mathcal{H}]\, dP = \int_H X\, dP \quad \forall H \in \mathcal{H} \qquad (2)$$

Let us establish the existence and uniqueness of $E[X|\mathcal{H}]$ by use of the Radon-Nikodym Theorem. Let ν_X be a measure on $\mathcal{H}$ defined as:

$$\nu_X(H) \triangleq \int_H X\, dP \quad \forall H \in \mathcal{H}$$

Then, $\nu_X << P|\mathcal{H} \Rightarrow$ there exists a $P|\mathcal{H}$-unique $\mathcal{H}$-measurable function $g_{X|\mathcal{H}}$ on Ω such that: $\nu_X(H) = \int_H g_{X|\mathcal{H}}\, dP \quad \forall H \in \mathcal{H}$

Therefore, $E[X|\mathcal{H}] \triangleq g_{X|\mathcal{H}}$ which is unique wrt the measure $P|\mathcal{H}$

Note that Eq (2) in Definition B.1 is equivalent to

$$\int_\Omega \langle Z, E[X|\mathcal{H}] \rangle\, dP = \int_\Omega Z\, dP \text{ for all } \mathcal{H}\text{-measurable fn } Z \qquad (3)$$

where $\langle \circ, * \rangle$ indicates the inner product of $\circ$ and $*$ in the usual sense.

Let $X : \Omega \to \mathbb{R}^n$ and $Y : \Omega \to \mathbb{R}^n$ two random vectors with $E[\|X_k\|] < \infty$ and $E[\|Y_k\|] < \infty \ \forall k$; let $a, b \in \mathbb{R}$ be constants.

(1) $E[(aX + bY)|\mathcal{H}] = a\,E[X|\mathcal{H}] + b\,E[Y|\mathcal{H}]$

(2) $E[E[X|\mathcal{H}]] = E[X]$

(3) $E[X|\mathcal{H}] = X$ if X is $\mathcal{H}$-measurable

(4) $E[X|\mathcal{H}] = X$ if X is independent of $\mathcal{H}$

(5) $E[\langle Y, X \rangle | \mathcal{H}] = \langle Y, E[X|\mathcal{H}] \rangle$ if Y is $\mathcal{H}$-measurable where $\langle Y, X \rangle$ is the inner product on $\mathbb{R}^n$ in the usual sense.

Proof: (1) follows directly from the linear property of the expectation operator,

(2) $E[E[X|\mathcal{H}]] = \int_\Omega E[X|\mathcal{H}] \, dP|\mathcal{H} = \int_\Omega X \, dP = E[X]$

(3) $E[X|\mathcal{H}] = \int_\Omega X \, dP|\mathcal{H}$ and X is $\mathcal{H}$-measurable

$$\Rightarrow E[X|\mathcal{H}] = \int_\Omega X \, dP = E[X]$$

(4) Since X is independent of $\mathcal{H}$, we have

$$E[X|\mathcal{H}] = \int_\Omega X \, dP|\mathcal{H} = \int_\Omega X \, dP = E[X]$$

(5) Let $G, H \in \mathcal{H}$. Define $\chi_H(\xi) = \begin{cases} 1 & \text{if } \xi \in H \\ 0 & \text{if } \xi \notin H \end{cases}$

Now, $\int_G \chi_H E[X|\mathcal{H}] \, dP = \int_{G \cap H} E[X|\mathcal{H}] dP = \int_{G \cap H} X \, dP = \int_G \chi_H X \, dP$

Let $Y_j \equiv c_j \chi_j$ be a simple function, where $H_j \in \mathcal{H} \; \forall j$

The proof follows immediately. ∎

Let $\mathcal{G}$ and $\mathcal{H}$ be σ-algebras such $\mathcal{G} \subset \mathcal{H}$. Then,

$$E[X|\mathcal{G}] = E[E[X|\mathcal{H}]|\mathcal{G}]$$

Proof: Let $G \in \mathcal{G} \Rightarrow G \in \mathcal{H}$. Then,

$$\int_G E[X|\mathcal{H}] \, dP = \int_G X \, dP$$

$\Rightarrow$ By the uniqueness property $\int_\Omega E[X|\mathcal{H}] \, dP|\mathcal{G} = \int_\Omega X \, dP|\mathcal{G}$

$$\Rightarrow E[E[X|\mathcal{H}]|\mathcal{G}] = E[X|\mathcal{G}]$$

Jensen Inequality: If $\varphi : \mathbb{R} \to \mathbb{R}$ is convex (i.e., $\forall \lambda \in (0,1)$ $\forall x, y \in \mathbb{R}$ $\varphi(\lambda x + (1-\lambda) y) \leq \lambda \varphi(x) + (1-\lambda)\varphi(y)$) and $E[|X|] < \infty$, then

$$\varphi(E[X|\mathcal{H}]) \leq E[\varphi(X)|\mathcal{H}]$$

Corollary 1: $E[X|\mathcal{H}] \leq E[|X| \, |\mathcal{H}]$

Corollary 2: $|E[X|\mathcal{H}]|^2 \leq E[|X|^2|\mathcal{H}]$

Corollary 3: If $X_k \to X$ in $L_2(P)$, then $E[X_k|\mathcal{H}] \longrightarrow E[X|\mathcal{H}]$ in the usual sense