1 Mathematical fundamentals.

1.1 Topology.

Definition 1 *A set A in R^k is called an open set if $\forall a \in A$, there exists $\varepsilon > 0$ such that $B_\varepsilon(a) \subset A$ (where $B_\varepsilon(a) = \left\{x \in R^k : \|x - a\| < \varepsilon\right\}$).*

Theorem 1 *The union of open sets is an open set.*

Definition 2 *A set A in R^k is called a closed set if $R^k \backslash A$ is open.*

Definition 3 *A set A in R^k is called a bounded set if there is $M \in R^+$ such that*

$$\forall a \in A, \|a\| \leq M$$

Definition 4 *A set A in R^k is called a compact set if it is closed and bounded.*

Definition 5 *A set A in R^k is called a compact set if for any family of open sets $\mathcal{F}$ such that,*

$$A \subset \bigcup_{B \in \mathcal{F}} B$$

there is a finite subfamily $\mathcal{F}_s$ such that

$$A \subset \bigcup_{B \in \mathcal{F}_S} B.$$

Definition 6 *A set A in R^k is called a convex set if $\forall a, b \in A$, then*

$$\lambda a + (1 - \lambda) b \in A, \forall \lambda \in [0, 1].$$

.

Definition 7 *A function*

$$f : R^m \to R^n$$

is a relationship in $R^m x R^n$ *such that* $\forall r \in R^m$ *there exists a unique value* $\bar{r} \in R^m$ *such that* $r f \bar{r}$ *or* $f(r) = \bar{r}$.

Definition 8 *A function* $f : R^m \to R^n$ *is continuous if for any open set in* R^n, *say* N, *we have that* $f^{-1}(N)$ *is an open set in* R^m.

Definition 9 *A function* $f : R^m \to R^n$ *is continuous at* a_0 *if for any* $\varepsilon > 0$ *there exists* $\delta_\varepsilon > 0$ *such that* $\forall a \in R^m, |a - a_0| < \delta_\varepsilon$ *implies that* $|f(a) - f(a_0)| < \varepsilon$

Definition 10 *A function* $f : R^m \to R^n$ *is uniformly continuous in a set* A *if if for any* $\varepsilon > 0$ *there exists* $\delta_\varepsilon > 0$ *such that for any* $a_0 \in A$ $|a - a_0| < \delta_\varepsilon$ *implies that* $|f(a) - f(a_0)| < \varepsilon$.

Theorem 2 *A continuous function defined over a compact set is uniformly continous with respect to that set.*

Definition 11 ***Definition 12*** *Consider a sequence of functions* $\{f_k\}_{k=1}^{\infty}$ *we say that*

$$\lim_{k \to \infty} f_k = f \ if \ \forall a \in R^m, \lim_{k \to \infty} f_k(a) = f(a)$$

or, for any $a \in S$ *and* $\forall \varepsilon > 0$

$$\exists k_a \in N \text{ such that } \forall k > k_a, |f_k(a) - f(a)| < \varepsilon$$

Theorem 3 Definition 13 *Consider a sequence of functions* $\{f_k\}_{k=1}^{\infty}$ *defined in a space* S*, we say that this sequence converges to a function* f *uniformly if*

$$\forall \varepsilon > 0 \exists k_o \in N \text{ such that } \forall k > k_o, |f_k(a) - f(a)| < \varepsilon \ \forall a \in S$$

Theorem 4 *Consider a sequence of functions* $\{f_k\}_{k=1}^{\infty}$*, such that* $\{f_k\}_{k=1}^{\infty}$ *converges to* f *uniformly in* S*. If* f_k *is continuous at a point* $c \in S$ f *is continuous at* c*.*

Proof. Because $\{f_k\}_{k=1}^{\infty}$ converges to f at any point in S, we have that for any $\varepsilon > 0$

$$\exists k_c \in N \text{ such that } \forall k > k_c, |f_k(c) - f(c)| < \varepsilon/3$$

On the other hand, because of continuity of f_k at c, there exists $\delta_c > 0$ such that,

$$x \in S \text{ and } |x - c| < \delta_c \text{ then } |f_k(x) - f_k(c)| < \varepsilon/3$$

Observe now that,

$$|f(x) - f(c)| \leq |f(x) - f_k(x)| + |f_k(x) - f_k(c)| + |f_k(c) - f(c)|$$

and the first and last term in the right hand side of the last equation are smaller or equal to $\varepsilon/3$ because of uniform convergence, and the second term is smaller or equal than $\varepsilon/3$ because of continuity. Thus, for any $\varepsilon > 0$ there exists $\delta_c > 0$ such that,

$$x \in S \text{ and } |x - c| < \delta_c \text{ then } |f(x) - f(c)| < \varepsilon$$

and this proofs the continuity of f at point c. ■

Example 1 *Consider*

$$f_k(x) = \left\{ \begin{array}{l} \text{-1 } if\ x \in (-\infty, -k^{-1}] \\ \\ \frac{k^{-1}+x}{k^{-1}} - 1\ if\ x \in (-k^{-1}, k^{-1}) \\ \\ \text{1 } if\ x \in [k^{-1}, \infty) \end{array} \right\}$$

at the limit, $k \to \infty$, $f_k(x) \to f(x)$

$$f(x) = \left\{ \begin{array}{l} \text{-1 } if\ x \in (-\infty, 0) \\ \\ 0\ if\ x = 0 \\ \\ \text{1 } if\ x \in (0, \infty) \end{array} \right\}$$

and $f(x)$ *is not continuos.*

Theorem 5 *Let* $f(\bullet)$ *be a continuous function defined in a compact set* A *then*

$$\exists \quad \underset{a \in A}{Max}\ f(a)$$

1.2 Probability Theory.

Let Ω be an space or set of points ω. From a probabilistic perspective a subset of Ω is an *event* and an element ω of Ω is a *sample point.*

Consider the experiment of tossing a coin. The space of events is $\Omega = \{tail, head\}$, the family of events is $\{\{tail\}, \{head\}, \{tail, head\}, \{\phi\}\}$.

The event $\{tail, head\}$ is the event that occurs if either $\{head\}$ or $\{tail\}$ occurs. It is in fact the union of the events $\{head\}$ and $\{tail\}$.

The event tail and head are complementary. That is, if $\{head\}$ occurs then $\{tail\}$ will not occur.

Intuitively, one would like to work with classes of events that have the property that given two events then the union of these events is also an event within the class of interest and the same with complements. The next definition formalizes this idea:

Definition 14 *Let Ω be an arbitrary nonempty space. A class $\mathcal{F}$ of subsets of Ω is called a* field (or algebra) *if it contains Ω itself and is closed under the formation of complements and finite unions:*

(i) $\Omega \in \mathcal{F}$

(ii) $A \in \mathcal{F}$ implies $A^c \in \mathcal{F}$

(iii) $A, B \in \mathcal{F}$ implies $A \cup B \in \mathcal{F}$

The previous definition implies that $\phi \in \mathcal{F}$. De Morgan's law, $A \cap B = (A^c \cup B^c)^c$ ...

Note that *(iii)* can be replaced by

(iii) $A, B \in \mathcal{F}$ implies $A \cap B \in \mathcal{F}$.

Definition 15 *Let Ω be an arbitrary nonempty space. A class $\mathcal{F}$ of subsets of Ω is called a* $\sigma-$field *if it is a field and if it is also closed under the formation of countable unions:*

(iv) $A_1, A_2, ... \in \mathcal{F}$ implies $A_1 \cup A_2 \cup ... \in \mathcal{F}$

Also, observe that

$$(A_1 \cup A_2 \cup ...)^c = A_1 \cap A_2 \cap .. \in \mathcal{F}.$$

The largest $\sigma-$*field* in Ω consists of all the subsers of Ω; the smallest $\sigma-$*field* in Ω consists of only ϕ and Ω.

Definition 16 *A set in $\mathcal{F}$ is called $\mathcal{F}-$*measurable.

Definition 17 *A pair $(\Omega, \mathcal{F})$ is a measurable space if $\mathcal{F}$ is a $\sigma-$*field *in Ω.*

The next definition presents a very important concept,

Definition 18 *Given a class of sets $\mathcal{A}$, we define the $\sigma-$*field *generated by $\mathcal{A}$ as the smallest $\sigma-$*field *containing $\mathcal{A}$ (or the intersection of all $\sigma-$*fields *containing $\mathcal{A}$) and is denoted $\sigma(\mathcal{A})$.*

If Ω is a topologycal space and B is the class of open sets in Ω then, $\sigma(B)$ is called Borel $\sigma-$*field* in Ω.

1.3 Probability Measure.

Definition 19 *A set function is a real-valued function defined on some class of subsets of Ω.*

Definition 20 *A set function P on a field $\mathcal{F}$ is a* measure *if it satisfies the following conditions*

(i) $P(A) \geq 0, \forall A \in \mathcal{F}$

$(ii) P(\phi) = 0.$

(iii) *If* $A_1, A_2, \ldots$ *is a disjoint sequence of* $\mathcal{F} -$ *sets*

and if $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$ *then,* $P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k).$

Definition 21 *A set function* P *on a field* $\mathcal{F}$ *is a* probability measure *if it is a measure and*

(iv) $0 \leq P(A) \leq 1, \forall A \in \mathcal{F}$ *and* $P(\Omega) = 1.$

Example 2 *Consider the experiment of tosing a coin. In this framework*

$$\begin{aligned} \Omega &= \{head, tail\} \\ \mathcal{F} &= \{\{head, tail\}, \{head\}, \{tail\}, \{\emptyset\}\} \\ P(\{head\}) &= p, P(\{\emptyset\}) = 0, \ldots \end{aligned}$$

Example 3 *Consider the set* R *of real numbers and consider*

$$\mathcal{A} = \{(a, b] : a, b \in R\}$$

a class of sets in R. Define the following set function in $\mathcal{A}$,

$$\lambda((a, b]) = b - a.$$

Then, we can use λ to define a measure in $\sigma(\mathcal{A})$. This measure is called the *Lebesge measure*. Note that $\sigma(\mathcal{A})$ is also the Borel measure defined with respect to the usual topology in R.

The extension from $\lambda(\bullet)$ defined in $\mathcal{A}$ to $\lambda(\bullet)$ defined in $\sigma(\mathcal{A})$ can be accomplished as a result of the following theorem:

Theorem 6 *Given $\mathcal{A}$ and $\lambda(\bullet)$ then,*

(i) If $B \in \sigma(\mathcal{A})$ and $\varepsilon > 0$,there exists a finite or infinite sequence $A_1, A_2, ..$ of sets in $\mathcal{A}$ such that $B \subset \bigcup_k A_k$ and $\mu\left(\left(\bigcup_k A_k\right) - B\right) < \varepsilon$.

(ii) If $B \in \sigma(\mathcal{A})$ and $\varepsilon > 0$, and if $\mu(B) < \infty$, then there exists a finite disjoint sequence $A_1, A_2, ..$ of sets in $\mathcal{A}$ such that $\mu\left(B \vartriangle \left(\bigcup_k A_k\right)\right) < \varepsilon$.

The concept of Lebesge measure can be extended to R^k. It is enough to define

$$\mathcal{A} = \left[\left\{x \in R^k : a_i < x_i \leq b_i, i = 1, ..., k\right\} : a, b \in R\right]$$

and,

$$\lambda_k\left((a, b]\right) = \prod_{i=1}^{k} (b_i - a_i).$$

Definition 22 *If P is a probability measure in $(\Omega, \mathcal{F})$ then $(\Omega, \mathcal{F}, P)$ is called a probability measure space.*

1.4 Measurable functions and random variables.

Definition 23 *Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be two measurable spaces. A function*

$$f : \Omega \to \Omega'$$

is measurable if $\forall A' \in \mathcal{F}', f^{-1}(A') \in \mathcal{F}$.

A continuous function between two measurable spaces is measurable with respect to the Borel $\sigma - field$.

Theorem 7 *Consider* $(\Omega, \mathcal{F}), (\Omega', \mathcal{F}')$ *and* $(\Omega'', \mathcal{F}'')$ *measurable spaces, and*

$$\begin{aligned} f &: \Omega \to \Omega' \\ g &: \Omega' \to \Omega'' \end{aligned}$$

measurable functions, then $g \circ f$ *is a measurable function.*

Corollary 8 *If* $f_i : \Omega \to R$ $i = 1, ...k$ *is measurable and* $g : R^k \to R$ *is measurable then* $g(f_1(\omega), ..., f_k(\omega))$ *is measurable. This is true in particular if* f_i $i = 1, ...k$ *and* g *are continuous.*

Definition 24 *A measurable, real-valued, function on a probability space* $(\Omega, \mathcal{F}, P)$ *is called a* random variable*.*

$$X : \Omega \longrightarrow R$$

Definition 25 *The distribution of a random variable* X *is an nonnegative function in* R

$$F_x : R \longrightarrow R_+,$$

such that

$$F_x(z) = P(X \leq z) = P(\omega \mid X(\omega) \leq z), \;\; z \in R.$$

As a result of the previous definition we have that $F(\bullet)$ is nondecreasing and rigth-continuous. Furthermore,

$$\begin{aligned}\lim_{x\longrightarrow-\infty} F(x) &= 0\\ \lim_{x\longrightarrow+\infty} F(x) &= 1.\end{aligned}$$

It is simple to show that $F_x(\bullet)$ is right-continous. Observe that

$$\lim_{\varepsilon\longrightarrow\mathbf{0}^+} F_x(z+\varepsilon) = \lim_{\varepsilon\longrightarrow\mathbf{0}} P(w \mid x(w) \leq z+\varepsilon) = P(w \mid x(w) \leq z) = F_x(z).$$

However,

$$\lim_{\varepsilon\longrightarrow\mathbf{0}^-} F_x(z+\varepsilon) = P(w \mid x(w) < z)$$

which is not necessarily equal to $F_x(z)$ and we cannot show that $F_x(\bullet)$ is left-continous.

A random variable and its distribution have density f with respect to the Lebesgue measure if f is a nonnegative function of R and

$$F(b) - F(a) = \int_a^b f(x)\,dx,$$

for $a, b \in R$, $a < b$.

1.5 Modes of Convergence.

Definition 26 *A sequence of random variables* $\{X_n\}_{n=1}^{\infty}$ *converges to a random variable* X in probability *if*

$$\lim_{n\longrightarrow\infty} P(|X_n - X| > \varepsilon) = 0.$$

We write

$$plim\ X_n = X \ or \ X_n \longrightarrow^p X.$$

Example 4 *Consider* $\{X_n\}_{n=1}^{\infty}$, *with* X_n *defined as*

$$X_n\,(head) = 1 + 1/n$$

$$X_n\,(tail) = 0$$

$$X\,(head) = 1$$

$$X\,(tail) = 0$$

for n *large enough*

$$P\left(|X_n - X| > \varepsilon\right) = P\left(|X_n\,(head) - X\,(head)| > \varepsilon\right) = P\left(1/n > \varepsilon\right) = 0.$$

Thus,

$$X_n \rightarrow_p X.$$

Definition 27 *A sequence of random variables* $\{X_n\}_{n=1}^{\infty}$ *converges to a random variable* X almost surely *if*

$$P\left(\omega \mid \lim_{n \longrightarrow \infty} X_n\,(\omega) = X\,(\omega)\right) = 1.$$

We write

$$plim\ X_n = X \ or \ X_n \longrightarrow^{a.s} X.$$

Definition 28 *A sequence of random variables* $\{X_n\}_{n=1}^{\infty}$ *converges to a random variable* X in mean square *if*

$$\lim_{n \longrightarrow \infty} E\left((X_n - X)^2\right) = 0.$$

We write

$$X_n \longrightarrow^M X.$$

Definition 29 *A sequence of random variables* $\{X_n\}_{n=1}^{\infty}$ *converges to a random variable* X in distribution *if the distribution function* $F_n(\bullet)$ *of* X_n *converges to the distribtion function* $F(\bullet)$ *of* X *at every continuity point of* $F(\bullet)$. *We write,*

$$X_n \longrightarrow^D X.$$

The next diagram shows the relationship between the four modes of convergence previousy defined.

$$\begin{array}{ccccc} & & a.s. & & \\ & & \downarrow & & \\ M & \longrightarrow & P & \longrightarrow & d \end{array}$$

Theorem 9 (***Chebyshev's Inequality***) *Let* $r > 0$ *and* X *a random variable such that* $E(|X|^r) < \infty$. *Then, for all* $\varepsilon > 0$,

$$P(|X| \geq \varepsilon) \leq \frac{E(|X|^r)}{\varepsilon^r}$$

Proof. Observe that

$$E(|X|^r) = \int_{-\infty}^{\infty} |X|^r \, dF \geq \int_{\{x:|x|\geq\varepsilon\}} |X|^r \, dF \geq \varepsilon^r \int_{\{x:|x|\geq\varepsilon\}} dF$$

$$\geq \varepsilon^r \left[F(-\varepsilon) + (1 - F(\varepsilon))\right],$$

because $\{x : |x| \geq \varepsilon\} = \{x : x \geq \varepsilon\} \cup \{x : x \leq -\varepsilon\}$. Also, taking into account that

$$[F(-\varepsilon) + (1 - F(\varepsilon))] = P(|X| \geq \varepsilon)$$

we get

$$E(|X|^r) \geq \varepsilon^r P(|X| \geq \varepsilon),$$

which is equivalent to Chebyshev's inequality.

■

The following result is a direct implication of Chebyshev's inequality.

Corollary 10 *Consider* $\{X_n\}$ *a sequence of random variables. Then,* $E(X_n^2) \longrightarrow 0$ *implies that* $X_n \longrightarrow^p 0$.

This result indicates that convergence in mean square error implies convergence in probability. Other relations between different modes of convergence are not proved.

Theorem 11 (*Algebra of Probability Limits*) *Consider two random sequences* $\{X_n\}, \{Y_n\}$ *and let* $p\lim X_n$ *and* $p\lim Y_n$ *exist. Then*

$p\lim X_n \pm Y_n = p\lim X_n \pm\, p\lim Y_n$

$p\lim X_n Y_n = p\lim X_n \cdot\, p\lim Y_n$

$p\lim X_n / Y_n = p\lim X_n /\, p\lim Y_n$, *provided that* $p\lim Y_n \neq 0$.

1.6 Orders of Convergence.

Definition 30 *Let $\{X_n\}$ be a sequence of random variables and let $\{a_n\}$ be a sequence of positive real numbers. Then, we can write $X_n = o(a_n)$ if*

$$p\lim a_n^{-1} X_n = 0$$

and $X_n = O(a_n)$ if for any $\varepsilon > 0$ there exists M_ε such that

$$P\left(a_n^{-1} |X_n| \le M_\varepsilon\right) \ge 1 - \varepsilon$$

for all values of n.

Lemma 12 *Let $X_n = o(a_n)$ and $Y_n = o(b_n)$. Then,*

$$\begin{aligned} X_n Y_n &= o(a_n b_n) \\ |X_n|^q &= o(a_n^q), \ q > 0 \\ X_n + Y_n &= o(\max\{a_n, b_n\}) \end{aligned}$$

Let $X_n = O(a_n)$ and $Y_n = O(b_n)$. Then,

$$\begin{aligned} X_n Y_n &= O(a_n b_n) \\ |X_n|^q &= O(a_n^q), \ q > 0 \\ X_n + Y_n &= O(\max\{a_n, b_n\}) \end{aligned}$$

Let $X_n = o(a_n)$ and $Y_n = O(b_n)$. Then,

$$X_n Y_n = o(a_n b_n)$$

Observe that the previous definition applies in particular when $\{X_n\}$ is a sequence of real numbers.

1.7 Laws of Large Numbers and Central Limit Theorems.

Theorem 13 (*Kolmogorov LLN2*) *Let $\{X_n\}$ be i.i.d. (independent and identically distributed). Then a necessary and sufficient condition that $\bar{X}_n \longrightarrow_{a.s.} \mu$ is that $E(X_n) = \mu$.*

Theorem 14 (*Lindeberg-Levy CLT*) *Let $\{X_n\}$ be i.i.d. with $E(X_n) = \mu$ and $Var(X_n) = \sigma^2$. Then*

$$\sqrt{n}\frac{\left(\bar{X}_n - \mu\right)}{\sigma} \longrightarrow_d N(0,1).$$