STA347 Probability I

Gun Ho Jang

Update on September 9, 2018

Note: This note is prepared for STA347. There might be numerous fault arguments/statements/typos. If you spot one, please contact the instructor or you may look up references which may contain errors too.

Review of Required Mathematics Background

This section summarises the minimum knowledge required for the course. You are assumed to understand the content of this section.

Set Theory

Set theory requires an axiomatic system. ZFC (Zermelo-Fraenkel + axiom of choice) is a well-established and well-accepted axioms.

In this course, higher level of set theory is accepted. Roughly saying, a *set* is a collection of distinguishable objects. Each member contained in a set is called an *element*.

Example.

- $\mathbb{N} = \{1, 2, 3, \ldots\}$ is the set of all natural numbers
- $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$ is the set of all integers
- $\mathbb{Q} = \{a/b : a, b \in \mathbb{Z}\}$ is the set of all rational numbers
- $\mathbb{R} = \overline{\mathbb{Q}}$ is the set of all real numbers
- $\mathbb{C} = \{a + ib : a, b \in \mathbb{R}\}$ is the set of all complex numbers

More definitions and properties are follows.

- A set containing no elements is called an *emptyset*, denoted by $\emptyset$.

- A set B is a *subset* of A ($B \subset A$) if all elements in B are also elements in A, that is, for all $b \in B$, $b \in A$.
- Two sets A and B are the same if and only if $A \subset B$ and $B \subset A$.
- The *cardinality* of a set A is the size of the set A, denoted by $|A|$. If there is a one-to-one correspondence, the cardicalities of two sets are the same.

  Note that $|\mathbb{N}| = |\mathbb{Z}| = |\mathbb{Q}| < |\mathbb{R}| = |\mathbb{C}|$.
- The *intersection* of sets A and B is the set only containing all common elements between A and B, that is, $A \cap B = \{c : c \in A, c \in B\}$
- The *union* of sets A and B is the set only containing elements of A or B, that is, $A \cup B = \{c : c \in A \text{ or } c \in B\}$.
- The *difference* of B from A is the set only containing elements of A which are not in B, that is, $A - B$ or $A \backslash B = \{c : c \in A, c \notin B\}$
- The *symmetric difference* of A and B is the set only containing elements contained in either A or B but not at the same time, that is, $A \Delta B = \{c : (c \in A, c \notin B) \text{ or } (c \notin A, c \in B)\}$ or $A \Delta B = (A - B) \cup (B - A)$.
- The *complement* of A is the set of all elements which are not in A, that is, $A^c = \{c : c \notin A\}$

Definition 1. A set A is *finite* if and only if it contains finite number of elements. A set B is *countable* if and only if there exists a one-to-one correspondence $f : B \to \mathbb{N} = \{1, 2, \ldots\}$. A set C is *uncountable* if it is neither finite nor countable.

Exercise 1. Let $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$ be sets of natural numbers, integers, rational numbers and real numbers.

Show that all of the above are infinite. Determine whether each set is countable or uncountable.

Theorem. (a) $A - B = A \cap B^c$. (b) [de Morgan's law] $(A \cap B)^c = A^c \cup B^c$ and $(A \cup B)^c = A^c \cap B^c$.

Proof. (a) First of all, $x \in A - B \Longrightarrow x \in A, x \notin B \Longrightarrow x \in A, x \in B^c \Longrightarrow x \in A \cap B^c$ implies $A - B \subset A \cap B^c$. On the other hand, $x \in A \cap B^c \Longrightarrow x \in A, x \in B^c \Longrightarrow x \in A, x \notin B \Longrightarrow x \in A - B$ implies $A \cap B^c \subset A - B$. Hence $A - B = A \cap B^c$.

(b) Using the definitions, we can write

$$\begin{aligned}(A \cup B)^c &= \{x : x \notin A \cup B\} \\ &= \{x : x \notin A \text{ and } x \notin B\} \\ &= \{x : x \in A^c \text{ and } x \in B^c\} \\ &= \{x : x \in A^c \cap B^c\} \\ &= A^c \cap B^c.\end{aligned}$$

$\square$

Exercise 2. Show that $(A \cap B)^c = A^c \cup B^c$.

Exercise 3. Show that

(a) $A \subset B$ and $B \subset C$ implies $A \subset C$.

(b) $\emptyset \subset A$.

(c) $A \cup (B \cup C) = (A \cup B) \cup C$.

(d) $A \cap (B \cap C) = (A \cap B) \cap C$.

(e) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

(f) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

For any sets A_i for $i \in I$ where I is any index set which could be finite, countable or uncountable,

$$\begin{aligned}\bigcap_{i \in I} A_i &= \{x : x \in A_i \text{ for all } i \in I\} \\ \bigcup_{i \in I} A_i &= \{x : x \in A_i \text{ for some } i \in I\}.\end{aligned}$$

When $I = \{1, 2\}$,

$$\begin{aligned}\bigcap_{i \in I} A_i &= A_1 \cap A_2 \\ \bigcup_{i \in I} A_i &= A_1 \cup A_2.\end{aligned}$$

Exercise. Show that

(a) $(\bigcup_{n=1}^{\infty} A_n)^c = \bigcap_{n=1}^{\infty} A_n^c$.

(b) $(A \cap \bigcup_{n=1}^{\infty} B_n)^c = A^c \cup \bigcap_{n=1}^{\infty} B_n^c$.

(c) $(A \cup \bigcup_{n=1}^{\infty} B_n)^c = A^c \cap \bigcap_{n=1}^{\infty} B_n^c$.

Sequences and Limits

- A *function* (or a mapping) f from a domain A to a codomain B ($f : A \to B$) is a relationship satisfying $f(a) \in B$ for all $a \in A$.

- A function f is *injective* if $f(a) \neq f(b)$ for all $a \neq b \in A$ and is *surjective* if for any $b \in B$ there exists an element $a \in A$ such that $f(a) = b$.

- A function f is $1-1$ *(one-to-one) correspondence* if f is both injective and surjective.

In this course we are concerned about real-valued or real-vector-valued functions unless the values are specified.

- A *sequence* x_n (or $\{x_n\}_{n=1}^{\infty}$) is an ordered collection of elements.

- A sequence x_n *converges* to x ($x_n \to x$) if and only if for any $\epsilon > 0$ there exists $N > 0$ such that $|x_n - x| < \epsilon$ for all $n \geq N$. Intuitively $x_n \to x$ if the difference $x_n - x$ diminishes to zero.

- A number x is called an *upper bound* of a set A if $a \leq x$ for any $a \in A$ and a *lower bound* of A if $a \geq x$ for all $a \in A$.

- The *supremum* of a set A is the *least upper bound*, that is, $\sup A = \sup(A) = x$ satisfying $a \leq x$ for all $a \in A$ and for any $y < x$ there exists $a \in A$ such that $a \geq y$.

- The *limit supremum* of a sequence x_n is defined by $\limsup_{n\to\infty} x_n = \lim_{m\to\infty} \sup_{n\geq m} x_n$.

- The *infimum* of a set A is the *greatest lower bound*, that is, $\inf A = \inf(A) = x$ satisfying $a \geq x$ for all $a \in A$ and for any $y > x$ there exists $a \in A$ such that $a \leq y$.

- The *limit infimum* of a sequence x_n is defined by $\liminf_{n\to\infty} x_n = \lim_{m\to\infty} \inf_{n\geq m} x_n$.

- A sequence x_n is said to be *Cauchy* if for any $\epsilon > 0$ there exists $N > 0$ such that $|x_n - x_m| < \epsilon$ for all $n, m \geq N$.

Theorem. A real-valued sequence x_n converges if it is Cauchy.

Theorem. A sequence x_n converges to x if and only if for any subsequence x_{n_k} there exists a further subsequence $x_{n_{k_l}}$ converging to x.

A function $f : \mathcal{X} \to \mathcal{Y}$

- is *continuous* at x if and only if $f(x_n)$ converges to $f(x)$ as $x_n \to x$ as $n \to \infty$ for any x,
- is *right continuous* at x if $f(x + h_n) \to f(x)$ for any x and $h_n \searrow 0$,
- is *left continuous* at x if $f(x + h_n) \to f(x)$ for any x and $h_n \nearrow 0$.

Exercise. Show that a function $f : \mathcal{X} \to \mathcal{Y}$ is continuous if and only if for any $x \in \Omega$ and $\epsilon > 0$ there exists $\delta > 0$ such that $|f(x) - f(z)| < \epsilon$ for any $z \in \Omega$ with $|z - x| < \delta$.

Exercise. A function f is continuous if and only if it is left and right continuous.

Exercise. Find a function that is right continuous but not left continous.

Measure Theory

Consider $\Omega = \mathbb{R}$, the length of an interval can be a size or a measurement of the interval. A measure of a subset $A \subset \mathbb{R}$ is an extended notion of the length like $b - a$ for an interval $A = (a, b]$.

Let μ be a measure on $\mathbb{R}$ extending the length, that is, $\mu : 2^{\mathbb{R}} \to [0, \infty]$. Naturally the measure μ is required to satisfy the following properties:

(a) $\mu((a, b]) = b - a$ for any $a < b \in \mathbb{R}$

(b) $\mu(\{a\}) = 0$ and $\mu(\emptyset) = 0$

(c) $\mu(A \cup B) = \mu(A) + \mu(B)$ for two disjoint sets A and B

(d) $\mu(\bigcup\limits_n B_n) = \sum\limits_n \mu(B_n)$ for mutually disjoint sets B_n

(e) $\mu(A^c)$ is defined as long as $\mu(A)$ is defined

(f) μ should be defined for any subset $A \subset \mathbb{R}$.

Concept behind (d), by considering $(0, 1] = \bigcup\limits_{n=1}^{\infty} (\frac{1}{n+1}, \frac{1}{n}]$, we want $1 = \mu((0, 1]) = \sum_{n=1}^{\infty} \mu((1/(n + 1), 1/n]) = \sum_{n=1}^{\infty} (1/n - 1/(n + 1)) = 1$.

Unfortunately, there is no function μ satisfying all (a)-(f). Mathematicians dropped (f) and restricted the domain of μ to $\mathcal{F}$ which is a collection of *measurable* sets.

The collection $\mathcal{F}$ satisfies

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

(b) for $A \in \mathcal{F}$, the complement $A^c \in \mathcal{F}$

(c) for $A_n \in \mathcal{F}$, the union $\bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{F}$.

Any collection $\mathcal{F}$ of Ω satisfying (a)-(c) is called a *σ-field*. Each element in $\mathcal{F}$ is called an *event*.

A mapping $\mu : \mathcal{F} \to [0, \infty]$ is called a *measure* if it satisfies

(a) $\mu(\emptyset) = 0$

(b) for a (countable) sequence of disjoint sets $A_n \in \mathcal{F}$, $\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$.

Note. Measures are extended notions of length, area and volumes.

Integral - Riemann Integral

Let f be a function defined on $[a, b]$. If Riemann sums, for $a = x_0 \le t_1 \le x_1 \le t_2 \le x_2 \le \cdots \le x_{n-1} \le t_n \le x_n = b$,

$$\sum_{i=1}^{n} f(t_i)(x_i - x_{i-1})$$

converges to a number S as the partition goes finer, that is, $\max(|x_i - x_{i-1}|, i = 1, \ldots, n) \to 0$ as $n \to \infty$, then the Riemann integeral of f on $[a, b]$ is S, that is,

$$\int_a^b f(x)\, dx = \lim_{n \to \infty} \sum_{i=1}^{n} f(t_i) \Delta_i.$$

Riemann integral is defined well for continuous functions and functions having countably many discontinuity points.

Integral - Riemann-Stieltjes integral

Let f, g be two functions defined on $[a, b]$. For $a = x_0 \le t_1 \le x_1 \le t_2 \le x_2 \le \cdots \le x_{n-1} \le t_n \le x_n = b$, consider the sum

$$\sum_{i=1}^{n} f(t_i)(g(x_i) - g(x_{i-1}))$$

which generalizes Riemann sum. If the above defined sum converges to S as the partition goes finer, then the Riemann-Stieltjes integeral of f with respect to g on $[a, b]$ is S, that is,

$$\int_a^b f(x)\, dg(x) = \lim_{n \to \infty} \sum_{i=1}^{n} f(t_i)(g(x_i) - g(x_{i-1})).$$

Riemann-Stieltjes integral is well-defined for continuous functions having at most countably many discontinuity points.

If g is continuous and differentiable, then

$$\int_a^b f(x)\ dg(x) = \int f(x)g'(x)\ dx.$$

If μ is a measure defined on $[a,b]$, then the Riemann-Stieltjes integral with respect to $g(x) = \mu([a,x])$ becomes

$$\int_a^b f\ d\mu = \int_a^b f(x)\ dg(x).$$

Limit of Integral

Suppose a sequence of integrable functions f_n are converging to an integrable function f.

Monotone Convergence Theorem. If $0 \leq f_1 \leq f_2 \leq \cdots$, $f_n \to f$ and $\int f(x)\ dx < \infty$, then

$$\int f_n(x)\ dx \to \int f(x)\ dx.$$

Fatou's Lemma. If $f_n(x) \geq 0$, then

$$\int \liminf_{n\to\infty} f_n(x)\ dx \leq \liminf_{n\to\infty} \int f_n(x)\ dx.$$

Dominated Convergence Theorem. If $f_n \to f$, $|f_n| \leq g$ and $\int g(x)\ dx < \infty$, then $|\int f(x)\ dx| < \infty$ and

$$\int f_n(x)\ dx \to \int f(x)\ dx.$$

Three convergence theorems are equivalent.

MCT implies Fatou's lemma: Let $g_n(x) = \inf_{m\geq n} f_m(x)$ so that $g_n(x) \leq f_m(x)$ for any $m \geq n$. The monotonicity implies $\int g_n(x)\ dx \leq \int f_m(x)\ dx$. By taking infimum,

$$\int g_n(x)\ dx \leq \inf_{m\geq n} \int f_m(x)\ dx.$$

By sending n to infinity, we get

$$\int \liminf_{n\to\infty} f_n(x)\ dx = \int \lim_{n\to\infty} \inf_{m\geq n} f_m(x)\ dx = \int \lim_{n\to\infty} g_n(x)\ dx \stackrel{\text{MCT}}{=} \lim_{n\to\infty} \int g_n(x)\ dx$$
$$\leq \lim_{n\to\infty} \inf_{m\geq n} \int f_m(x)\ dx = \liminf_{n\to\infty} \int f_n(x)\ dx.$$

Fatou's lemma implies DCT: From the facts $|f_n(x)| \leq g(x)$ and $f_n(x) \to f(x)$, we get $|f(x)| \leq g(x)$. Hence $|\int f(x)\ dx| \leq \int |f(x)|\ dx \leq \int g(x)\ dx$.

Note that $f_n(x)+g(x) \geq 0$, $|f_n(x)+g(x)| \leq |f_n(x)|+g(x) \leq 2g(x)$, and $f_n(x)+g(x) \to f(x)+g(x) \leq 2g(x)$ and

$$\int f(x) + g(x)\ dx = \int \liminf_{n\to\infty} (f_n(x) + g(x))\ dx \leq \liminf_{n\to\infty} \int f_n(x) + g(x)\ dx = \liminf_{n\to\infty} \int f_n(x)\ dx + \int g(x)\ dx$$

By subtracting $0 \leq \int g(x)\ dx < \infty$, we get

$$\int f(x)\ dx \leq \liminf_{n\to\infty} \int f_n(x)\ dx.$$

Similarly, $g(x) - f_n(x) \geq 0$ and $g(x) - f_n(x) \to g(x) - f(x)$ and

$$\int g(x)\ dx - \int f(x)\ dx \leq \liminf_{n\to\infty} \Big[\int g(x)\ dx - \int f_n(x)\ dx \Big] = \int g(x)\ dx - \limsup_{n\to\infty} \int f_n(x)\ dx$$

which implies

$$\int f(x) \geq \limsup_{n\to\infty} \int f_n(x)\ dx$$

Finally

$$\int f(x) \ \leq \liminf_{n\to\infty} \int f_n(x)\ dx \leq \limsup_{n\to\infty} \int f_n(x)\ dx \leq \int f(x)\ dx$$

implies $\int f(x)\ dx = \lim_{n\to\infty} \int f_n(x)\ dx$.

DCT implies MCT: Note that $0 \leq f_n(x) \leq f(x)$ and $\int f(x)\ dx < \infty$. DCT directly implies $\int f_n(x)\ dx \to \int f(x)\ dx$.