Measure Theory, Limit Theorems, and Random Variables

Lee De Zhang

December 2, 2020

Contents

1 Introduction

This chapter outlines the tools necessary for a more rigorous study of probability.

2 Measure Theory

An algebra $\mathcal{F}$ is said to be a σ-algebra if

1. It is closed under complement, i.e., $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$.
2. Closed under countable union, i.e., $A_i \in \mathcal{F} \Rightarrow \cup_i A_i \in \mathcal{F}$ for countable number of A_i.

Remark 1. *Countable refers to finite or countably infinite. Some textbooks also state closed under countable intersection, but it is easy to derive that from the above conditions.*

.

We will now state the definition of a probability space

Definition 2.1 (*Probability Space*)**.** *A probability space is the triple $(\Omega, \mathcal{F}, P)$, such that Ω is the support or set of all possible outcomes (sample space), $\mathcal{F}$ is a set of events, and $P : \mathcal{F} \to [0, 1]$ which assigns probabilities to items in $\mathcal{F}$. $\mathcal{F}$ is a σ-algebra, which is a non-empty collection of subsets of Ω and satisfies the properties above.*

Without P, the double $(\Omega, \mathcal{F})$ is a measurable space, i.e., something which we can assign a measure to. We will define a measure next.

Definition 2.2 (*Measure*). *A measure is a non-negative countably additive set function, such that given a σ-algebra $\mathcal{F}$ and a measure μ, then $\mu : \mathcal{F} \to \mathbb{R}$ such that*

1. $\mu(A) \geq \mu(\emptyset) = 0 \quad \forall A in \mathcal{F}$.
2. *For $i = 1, 2, \cdots$, if $A_i \in \mathcal{F}$ is a countable sequence of disjoint sets, then $\mu(\cup A_i) = \sum \mu(A_i)$.*

Remark 2. *If $\mu(\Omega) = 1$, then μ is a probability measure.*

.

From the above definition, we can deduce that a measure has the following properties.

Theorem 2.3 (*Properties of a measure*). *Suppose μ is some measure on a measurable space $(\Omega, \mathcal{F})$, then μ satisfies,*

1. *Monotonicity. $A \subset B \Rightarrow P(B) - P(A) - P(B - A) \geq 0$ for $A, B \in \mathcal{F}$.*
2. *Subadditivity. For $m = 1, \cdots$, if $A_m \in \mathcal{F}, A \subset \cup_{m=1}^{\infty} A_m$, then $P(A) \leq \sum_{m=1}^{\infty} P(A_m)$.*
3. *Continuity from below. If $A_i \uparrow A$, i.e, $A_1 \subset A_2 \subset \cdots$ and $\cup A_i = A$, then $P(A_i) \uparrow P(A)$.*
4. *Continuity from above. If $A_i \downarrow A$, then $P(A_i) \downarrow P(A)$.*

Proof. **Monotonicity**. Since $B - A = B \cap A^c$ and A and $B - A$ are disjoint, then $\mu(A) = \mu(B) + \mu(B - A)$ and the result follows from the the non-negativity of μ.

Subadditivity. Let $B_1 = A_1$ and $B_k = A_k - \cup_{i=1}^{k-1} A_i$, then $\mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(B_i)$, the result follows from monotonicity.

Continuity from below. Let $B_k = A_k - A_{k-1}$. Then $\mu(A_k) = \sum_{i=1}^{k} \mu(B_i)$. Then the result follows.

Continuity from above. Observe that $A_i^c \uparrow A$, then $P(A_i^c) \uparrow P(A^c)$ and the result follows. $\square$

To make things a bit more concrete, let us apply this to an example on the discrete probability space.

Example 2.1 (*Discrete Probability Spaces*). *Let Ω be a countable set, $\mathcal{F}$ be the set of all subsets of Ω, for $A \in \mathcal{F}$, $P(A) = \sum (x \in A) p(x) \geq 0$ and $\sum_{\omega \in \Omega} p(\omega) = 1$. If Ω is finite, then $p(\omega) = 1/|\Omega|$. For example, when rolling a fair dice, $\Omega = \{1, 2, 3, 4, 5, 6\}$.*

Example 2.2 (*Measure on continuous real line*). *Let $\mathbb{R}$ be the real line, $\mathcal{R}$ be the Boreal sets (smallest σ-algebra containing the open sets), λ be the Lebesgue measure (for $a < b$, $\lambda((a, b]) = b - a$, and $\lambda(\mathbb{R}) = \infty$. To get a probability space, let $\Omega = (0, 1)$, $\mathcal{F} = \{A \cap (0, 1) : A \in \mathcal{R}\}$, and $P = \lambda$. Then P is a Lebesgue measure restricted to the Borel subsets in interval (0,1).*

Theorem 2.4 (*σ-algebra closed under arbitrary intersection*). *If $\mathcal{F}_i, i \in I$ for an arbitrary (possibly uncountable) I, then $\cap_i \mathcal{F}_i$ is a σ-algebra.*

Proof. Since $\emptyset, \Omega \in \mathcal{F}_i$, then $\emptyset, \Omega \in \cap\mathcal{F}_\rangle$. Furthermore, $A \in \cap\mathcal{F}_i \Rightarrow A \in \mathcal{F}_i$ hence $\cap\mathcal{F}_i$ closed under intersection and complement. □

Definition 2.5 (*Smallest σ-field*)**.** *Let $\mathcal{A}$ be a collection of subsets of Ω. Then there is a smallest σ-field containing $\mathcal{A}$, written as $\sigma(\mathcal{A})$.*

Remark 3. *The existence of $\sigma(\mathcal{A})$ is guaranteed using Theorem 2.4. Namely, let F_i be the different σ-algebra that contain $\mathcal{A}$. Then an intersection yields the smallest one.*

Theorem 2.6 (*Measure on Product Spaces*)**.** *If $(\Omega_i, \mathcal{F}_i, P_i), i = 1, \cdots, n$ are probability spaces, let $\Omega = \Omega_1 \times \cdots \times \Omega_n = \{(\omega_1, \cdots, \omega_n) : \omega_i \in \Omega_i\}$, $\mathcal{F} = \prod_{i=1}^n \mathcal{F}_i$ is the σ-field generated by $\{A_1 \times \cdots \times A_n : A_i \in \mathcal{F}_i\}$. Then P is a measure on $\mathcal{F}$, given by,*

$$P(A_1 \times \cdots \times A_n) = \prod_{i=1}^{n} P_i(A_i).$$

We can define a random variable on probability spaces. A real valued function X defined on Ω is a random variable if for every borel set $B \subset \mathbb{R}$,

$$X^{-1}(B) = \{\omega : X(\omega) \in B\} \in \mathcal{F}.$$

If X is a random variable, then X induces a probability measure on $\mathbb{R}$ called its distribution, by setting $\mu(A) = P(X \in A) = P(X^{-1}(A)) = P(\{\omega : X(\omega) \in A\})$ for a given Borel set $A \in \mathbb{R}$. Equivalently, this means taking $X^{-1}(A) \in \mathcal{F}$ and measuring that set. In this case, we say that X is $\mathcal{F}$ measurable ($X \in \mathcal{F}$). Furthermore, if Ω is discrete, then $X : \Omega \to \mathbb{R}$.

It is straightforward to verify that μ is a probability measure on $(\mathbb{R}, \mathcal{R}, \mu)$. Let $A_i \in \mathcal{R}$ be arbitrary disjoint sets. Then $X \in \cup A_i$ iff X falls in exactly one of the A_i. Therefore,

$$\mu(\cup A_i) = P(X \in \cup A_i) = P(\cup\{X \in A_i\}) = \sum P(X \in A_i) = \sum \mu(A_i).$$

An alternative way of thinking of random variables is that, suppose X is a random variable, then $X \in A = \{\omega : X(\omega) \in A\}$ is whether the realization of X falls in the events contained in A.

The distribution of X can be described using its distribution function $F(x) = P(X \leq x)$, which has the following properties.

Theorem 2.7 (*Properties of Distribution Function*)**.** *All distribution functions F have the following properties,*

1. *F is non-decreasing.*
2. *$\lim_{x\to\infty} F(x) = 1, \lim_{x\to-\infty} F(x) = 0$.*
3. *F is right continuous, $\lim_{y\downarrow x} F(y) = F(x)$.*
4. *$F(x-) = \lim_{y\uparrow x} F(y) = P(X > x)$.*
5. *$P(X = x) = F(x) - F(X-)$.*

Proof. 1. Suppose $x < y$ then $\{X < x\} \subset \{X < y\}$ therefore $F(x) \leq F(y)$.

2. $\lim x \to \infty \{X < x\} = \Omega$ and $\lim x \to -\infty \{X < x\} = \emptyset$.

3. $\lim_{y \downarrow x} \{X < y\} = \{X < x\}$.

4. $\lim_{y \uparrow x} \{X < y\} = \{X < x\}$.

5. Since $\{X = x\} = \{X \geq x\} \cap \{X < x\}^c$, then taking limits on both sides using 3, 4 gives the result.

□

Corollary 2.8. *If F satisfies 1, 2, 3 of Theorem 2.7, then it is a distribution function of some random variable.*

Proof. We prove this by constructing the random variable.

Let $\Omega = (0, 1)$, $\mathcal{F} =$ borel sets, $P =$ Lebesgue measure. For some $\omega \in [0, 1]$, let $X(\omega) = \sup\{y : F(y) < w\}$. Then, $\{\omega : X(\omega) < x\} = \{\omega : \omega < F(x)\}$, then $P(\omega : \omega \leq F(x)) = F(x)$ such that P is the Lebesgue measure. The result then follows. □

The above result is used to generate random variables from a uniform random variable.

Corollary 2.9. *Furthermore, if F satisfies 1, 2, 3 of Theorem 2.7, then there is a unique probability measure μ on $(\mathbb{R}, \mathcal{R})$ such that $\mu((a, b]) = F(b) - F(a)$ for $b > a$.*

Definition 2.10 (*Equal in Distribution*)**.** *Given two random variables X and Y on a common probability space. If both induce the same distribution μ on the probability space, i.e. for any x, $P(X < x) = P(Y < x)$, then we say both are equal in distribution, written as $X \stackrel{d}{=} Y$.*

Finally, the density function f of a random variable X is given by,

$$P(X = x) = \lim_{\epsilon \to 0} \int_{x-\epsilon}^{x+\epsilon} f(y)\, dy.$$

and its relationship with F is,

$$F(x) = \int_{-\infty}^{x} f(x)\, dx.$$

Example 2.3 (*Uniform Distribution on Cantor Set*)**.** *The cantor set C is defined by removing $(1/3, 2/3)$ from $[0, 1]$, and recursively removing the middle third of each interval that remains. Suppose we assign the distribution function such that*

- *$F(0) = 0$,*
- *$F(x) = 1/2$ for $x \in [1/3, 2/3]$,*
- *$F(x) = 1/4$ for $x \in [1/9, 2/9]$.*
- *and so on...*

The F that results is known as Lebesgue's singular function, because there is no f such that $F(x) = \int_{-\infty}^{x} f(x)\, dx$. It is also immediate that $\mu(C^c) = 0$.

Definition 2.11 (*Discrete Probability Measure*)**.** *A probability measure P is said to be discrete if there is a countable set S such that $P(S^c) = 0$.*

We list an interesting example of a discrete probability measure

Example 2.4 (*Dense discontinuities*)**.** *Let $q_1, q_2, \cdots$ be an enumeration of all the rationals, and*

$$F(x) = \sum_{i=1}^{\infty} \frac{1}{2^i} 1_{[q_i,\infty)}.$$

Then, let,

$$S = \{q_i : i = 1, 2, \cdots\},$$

such that $S^c = \mathbb{R} - \mathbb{Q}$. For any $x \in S^c$, $P(X = x) = 0$, hence X is discrete.

3 Random Variables

We proceed to prove that random variables are indeed measurable.

Definition 3.1 (*Measurable Map*)**.** *A function $X : \Omega \to S$ is a measurable map from $(\Omega, \mathcal{F})$ to $(S, \mathcal{S})$ if,*

$$X^{-1}(B) = \{\omega : X(\omega) \in B\} \in \mathcal{F},$$

for all $B \in \mathcal{S}$.

Theorem 3.2. *If $\{\omega : X(\omega) \in A\} \in \mathcal{F}$ for all $A \in \mathcal{A}$ and $S = \sigma(\mathcal{A})$, then X is measurable.*

Proof. We introduce the notation $\{X \in B\} = \{\omega : X(\omega) \in B\}$. Therefore, $\{X \in \cup B_i\} = \cup\{X \in B_i\}$ and $\{X \in B^c\} = \{X \in B\}^c$, so closure under complement and arbitrary union means $\mathcal{B} = \{B : \{X \in B\} \in \mathcal{F}\}$ is a σ-algebra. Since $\mathcal{A} \subset \mathcal{B}$ and $\mathcal{A} \subset \mathcal{S} = \sigma(\mathcal{A})$, then by definition, $\sigma(\mathcal{A}) = \mathcal{S} \subset \mathcal{B}$. $\square$

Corollary 3.3 (*σ-field generated by X*)**.** *If $\mathcal{S}$ is a σ-field, then $\{\{X \in B\} : B \in \mathcal{S}\}$ is a σ-field. It is the smallest σ field on Ω which makes X a measurable map, known as the σ-field generated by X, $\sigma(X)$.*

Theorem 3.4. *If $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ and $f : (S, \mathcal{S}) \to (T, \mathcal{T})$ are measurable maps, then $f(X)$ is a measurable map from $(\Omega, \mathcal{F})$ to $(T, \mathcal{T})$*

Proof. Let $B \in \mathcal{T}$ and hence $f^{-1}(B) \in \mathcal{S}$. Therefore

$$\{\omega : f(X(\omega)) \in B\} = \{\omega : X(\omega) \in f^{-1}(B)\}.$$

$\square$

Theorem 3.5. *If $X_1, \cdots, X_n$ are random variables on $(\Omega, \mathcal{F})$, $f : (\mathbb{R}^n, \mathcal{R}^n) \to (\mathbb{R}, \mathcal{R})$ is measurable, then $f(X_1, \cdots, X_n)$ is a random variable.*

Proof. Let A_i be Borel sets, then

$$\{(X_1, \cdots, X_n) \in A_1 \times \cdots \times A_n\} = \cap\{X_i \in A_i\} \in \mathcal{F},$$

and since $A_1 \times \cdots \times A_n$ generate $\mathcal{R}^n$ the result follows. $\square$

Theorem 3.6. *If $X_1, \cdots$ are random variables, then so are* $\inf X_n, \sup X_n, \limsup X_n, \liminf X_n$*.*

Proof. inf and sup can be proved similarly. We observe that

$$\{\sup X_n < x\} = \cap\{X_i < x\} \in \mathcal{F},$$

and

$$\liminf X_n = \sup_n \inf_{m>n} X_m,$$

and since $\inf_{m>n} X_m \in \mathcal{F}$ so $\sup_n \inf_{m>n} X_m \in \mathcal{F}$.

$\square$

Definition 3.7 (*Almost sure convergence*)**.** $X_n \overset{a.s.}{\to}$ *iff*

$$P\left(\{\omega : \liminf X_n = \limsup X_n\}\right),$$

that is, $\lim X_n$ *exists.*

4 Expected Value

Given a random variable $X \geq 0$ on $(\Omega, \mathcal{F}, P)$, then we define expectation as $E\,X = \int X\,dP$.

Definition 4.1 (x^+ *and* x^-)**.** *Define the positive part* $x^+ = \max(x, 0)$ *and* $x^- = \max(0, -x)$*.*

Remark 4. *$E\,X$ exists iff $E\,X = E\,X^+ - E\,X^-$ exists, i.e., both terms on right exist.*

Theorem 4.2. *Suppose $X, Y \geq 0$ or $E|X|, E|Y| < \infty$. Then,*

- $E(X + Y) = E(X) + E(Y)$
- $E(aX + b) = aE(X) + b$
- $X \geq Y$ *then* $E(X) \geq E(Y)$*.*

Theorem 4.3. *Suppose $E|X|, E|Y| < \infty$, then $EX = EY$ iff $X = Y$ a.s.*

Proof. Since $X \geq Y$ then $X - Y \geq 0$, So, $E|X - Y| = E(X - Y) = EX - EY = 0$ hence for all $\epsilon > 0, P(|X - Y| > \epsilon) = 0$. $\square$

Theorem 4.4 (*Fatou's Lemma*)**.** *If $X_n > 0$ then* $\liminf EX_n \geq E \liminf X_n$*.*

Theorem 4.5 (*Monotone Convergence Theorem*)**.** *If $0 \leq X_n \uparrow X$ then $EX_n \uparrow EX$.*