Chapter 1

Probability Theory: Introduction

Basic Probability – General

- In a *probability space* (Ω, Σ, P), the set Ω is the set of all possible *outcomes* of a "probability experiment". Mathematically, Ω is just a set, with elements ω. It is called the *sample space.*

- An *event* is the answer to a Yes/No question. Equivalently, an event is a subset of the probability space: A ∈ Ω. Think of A as the set of outcomes where the answer is "Yes", and A^c is the complementary set where the answer is "No".

- A *σ-algebra* is a mathematical model of a state of partial knowledge about the outcome. Informally, if Σ is a σ-algebra and A ∈ Ω , we say that A ∈ Σ if we know whether ω ∈ A or not.

Definitions – Algebra

• We want to put some structure in the sets (data) we work. We would like to be able to measures the sets and be able to easily manipulate them (algebraically).

Definitions: Algebra

A collection of sets F is called an *algebra* if it satisfies:

- $\emptyset \in F$.
- If $\omega_1 \in F$, then $\omega_1^C \in F$. (F is closed under *complementation*)
- If $\omega_1 \in F$ & $\omega_2 \in F$, then $\omega_1 \cup \omega_2 \in F$. ($F$ is closed under *finite unions*).

Note: The set $E = \{\{\Phi\},\{a\},\{b\},\{a,b\}\}$ is an algebra.

Definitions: Sigma-algebra

Definition: Sigma-algebra

A *sigma-algebra* (σ-algebra or σ-field) F is a set of subsets ω of Ω s.t.:

- $\emptyset \in F$.
- If $\omega \in F$, then $\omega^C \in F$. (ω^C = complement of ω)
- If $\omega_1, \omega_2, \ldots, \omega_n, \ldots \in F$, then, $\bigcup_{i=1}^{\infty} \omega_i \in F$ (ω_{1i}'s are *countable*)

Note: The set $E = \{\{\Phi\},\{a\},\{b\},\{c\},\{a,b\},\{b,c\},\{a,c\},\{a,b,c\}\}$ is an algebra and a σ-algebra.

σ-algebras are a subset of algebras in the sense that all σ-algebras are algebras, but not vice versa. Algebras only require that they be closed under pairwise unions while σ-algebras must be closed under *countably* infinite unions.

Sigma-algebra

Theorem:
All σ-algebras are algebras.

• Sigma algebras can be generated from arbitrary sets. This will be useful in developing the probability space.

Theorem:
For some set X, the intersection of all σ-algebras, A_i, containing X −that is, $x \in X \Rightarrow x \in A_i$ for all i− is itself a σ-algebra, denoted σ(X).

$\Rightarrow$ This is called the *σ-algebra generated* by X.

Sample Space, Ω

Definition: Sample Space
The *sample space* **Ω** is the set of all possible unique outcomes of the experiment at hand.

Example: If we roll a die, **Ω** = {1; 2; 3; 4; 5; 6}.

In the probability space, the σ-algebra we use is σ(Ω), the σ-algebra generated by Ω. Thus, take the elements of Ω and generate the "extended set" consisting of all unions, compliments, compliments of unions, unions of compliments, etc. Include Φ*;* with this "extended set" and the result is σ (Ω), which we denote as **Σ**.

Definition The σ-algebra generated by Ω, denoted **Σ**, is the collection of possible events from the experiment at hand.

(Ω, Σ)

Definition The σ-algebra generated by Ω, denoted Σ, is the collection of possible events from the experiment at hand.

Example: We have an experiment with **Ω** = {1, 2}. Then,
Σ = {{Φ},{1},{2},{1,2}}. Each of the elements of Σ is an event. Think of events as descriptions of experiment outcomes (Φ: the "nothing occurs" event).

Note that σ-algebras can be defined over the real line as well as over abstract sets. To develop this notion, we need the concept of a *topology*.

Note: There are many definitions of topology based on the concepts of neighborhoods, open sets, closed set, etc. We present the definition based on open sets.

Topological Space

Definition (via open sets):
A *topological space* is an ordered pair (X, τ), where X is a set and τ is a collection of subsets of X, satisfying:

1. Ø; $X \in \tau$
2. τ is closed under finite intersections.
3. τ is closed under arbitrary unions.

Any element of a topology is known as an *open set*. The collection τ is called a *topology* on X.

Example: We have an experiment with Ω = {1, 2, 3}. Then,

τ = {{Ø}, {1,2,3}} is a (trivial) topology on Ω.
τ = {{Ø}, {1}, {1,2,3}} is also a topology on Ω.
τ = {{Ø}, {1,2}, {2,3}, {1,2,3}} is NOT a topology on Ω.

Borel σ-algebra

Definition: Borel σ-algebra (Emile Borel (1871-1956), France.)

The *Borel σ-algebra* (or, *Borel field*) denoted B, of the topological space (X; τ) is the σ-algebra generated by the family τ of open sets. Its elements are called Borel sets.

Lemma: Let C = $\{(a; b): a < b\}$. Then σ(C) = B_R is the Borel field generated by the family of all open intervals C.

What do elements of B_R look like? Take all possible open intervals. Take their compliments. Take arbitrary unions. Include Ø and R. B_R contains a wide range of intervals including open, closed, and half-open intervals. It also contains disjoint intervals such as $\{(2; 7]\ U\ (19; 32)\}$. It contains (nearly) every possible collection of intervals that are imagined.

Measures

Definition: Measurable Space

A pair (X, Σ) is a *measurable space* if X is a set and Σ is a nonempty σ-algebra of subsets of X.

A measurable space allows us to define a function that assigns real-numbered values to the abstract elements of Σ.

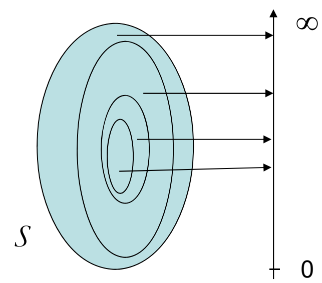
Definition: Measure μ

Let (X, Σ) be a measurable space. A set function μ defined on Σ is called a *measure* iff it has the following properties.

1. $0 \leq \mu(A) \leq \infty$ for any $A \in \Sigma$.
2. $\mu(\Phi) = 0$.
3. (σ-additivity). For any sequence of pairwise disjoint sets $\{A_n\} \in \Sigma$ such that $\text{U}_{n=1}\, A_n \in \Sigma$, we have $\mu(U_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$

Measures

Intuition: A measure on a set, S, is a systematic way to assign a positive number to each suitable subset of that set, intuitively interpreted as its *size*. In this sense, it generalizes the concepts of length, area, volume.



Examples (of measures):
- Counting measure: $\mu(S)$ = number of elements in S.
- Lebesgue measure on R: $\mu(S)$ = conventional length of S.
 That is, if $S = [a,b] \Rightarrow \mu(S) = \lambda[a,b] = b - a$.

Measures & Measure Space

• Note: A measure μ may take ∞ as its value. Rules:

(1) For any $x \in R$, $\infty + x = \infty$, $x * \infty = \infty$ if $x > 0$, $x * \infty = -\infty$ if $x < 0$, and $0 * \infty = 0$;

(2) $\infty + \infty = \infty$;

(3) $\infty * a = \infty$ for any $a > 0$;

(4) $\infty - \infty$ or ∞/∞ are not defined.

Definition: Measure Space

A triplet (X, Σ, μ) is a *measure* space if (X, Σ) is a measurable space and $\mu: \Sigma \rightarrow [0; \infty)$ is a measure.

• If $\mu(X) = 1$, then μ is a *probability measure*, which we usually use notation P, and the measure space is a *probability space*.

Lebesgue Measure

• There is a unique measure λ on (R, B_R) that satisfies

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

for every finite interval $[a, b]$, $-\infty < a \leq b < \infty$. This is called the *Lebesgue measure.*

If we restrict λ to the measurable space $([0, 1], B_{[0,1]})$, then λ is a *probability measure.*

Examples:

- Any Cartesian product of the intervals $[a, b]$ x $[c, d]$ is Lebesgue measurable, and its Lebesgue measure is $\lambda = (b - a)*(d - c)$.
- λ([set of rational numbers in an interval of R]) = 0.

Note: Not all sets are Lebesgue measurable. See Vitali sets.

Zero Measure

Definition: Measure Zero

A (μ-)measurable set E is said to have (μ-)*measure zero* if $\mu(E) = 0$.

Examples: The singleton points in R^n, and lines and curves in R^n, $n \geq 2$. By countable additivity, any countable set in R^n has measure zero.

• A particular property is said to hold *almost everywhere* ("*a.e.*") if the set of points for which the property fails to hold is a set of measure zero.

Example: "a function vanishes almost everywhere"; "$f = g$ a.e.".

• Note: Integrating anything over a set of measure zero produces 0. Changing a function on a set of measure zero does not affect the value of its integral.

Measure: Properties

• Let (Ω, Σ, μ) be a measure space. Then, μ has the following properties:

(i) (Monotonicity). If A ⊂ B, then μ(A) ≤ μ(B).

(ii) (Subadditivity). For any sequence $A_1; A_2; \ldots$

$$\mu[\cup_{i=1} A_i] \leq \sum_i^\infty \mu[A_i]$$

(iii) (Continuity). If $A_1 \subset A_2 \subset A_3 \subset$ .... (or $A_1 \supset A_2 \supset A_3 \supset$ ... and $\mu(A_1) < \infty$), then

$$\mu(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} \mu(A_n),$$

where

$$\lim_{n\to\infty} A_n = \cup_{i=1} A_i \qquad \text{(or } \cup_{i=1} A_i\text{)}$$

Probability Space

A measure space (Ω, Σ, μ) is called finite if μ(Ω) is a finite real number (not ∞). A measure μ is called *σ-finite* if Ω can be decomposed into a countable union of measurable sets of finite measure.

For example, the real numbers with the Lebesgue measure are σ-finite but not finite.

Definition: Probability Space

A measure space is a probability space if μ(Ω)=1. In this case, μ is a *probability measure,* which we denote P.

• Let P be a probability measure. The *cumulative distribution function* (*c.d.f.*) of P is defined as:

$$F(x) = P((-\infty, x]),\ x \in R$$

Measurable Function

Definition: Inverse function

Let f be a function from Ω to Λ (often $\Lambda = R^k$)

Let the inverse image of $B \subset \Lambda$ under f:

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

• Useful properties:

– $f^{-1}(B^c) = (f^{-1}(B))^c$ for any $B \subset \Lambda$;

– $f^{-1}(\cup B_i) = \cup f^{-1}(B_i)$ for any $B_i \subset \Lambda$, i = 1, 2, ...

Note: The inverse function f^{-1} need not exist for $f^{-1}(B)$ to be defined.

Definition: Let (Ω, Σ) and (Λ, G) be measurable spaces and f a function from Ω to Λ. The function f is called a *measurable function* from (Ω, Σ) to (Λ, G) if and only if $f^{-1}(G) \subset \Sigma$.

Measurable Function

• If f is measurable from (Ω, Σ) to (Λ, G) then $f^{-1}(G)$ is a sub-σ-field of Σ. It is called the σ-field generated by f and is denoted by $\sigma(f)$.

• If f is measurable from (Ω, Σ) to (R, B), it is called a *Borel function* or a *random variable* (RV).

• A random variable is a convenient way to express the elements of Ω as numbers rather than abstract elements of sets.

Measurable Function

Example: Indicator function for $A \subset \Omega$.

$$I_A(\omega) = 1 \quad \text{if } \omega \in A$$
$$= 0 \quad \text{if } \omega \in A^C$$

For any $B \subset R$

$$I_A^{-1}(B) = \emptyset \quad \text{if 0 not in B, 1 not in B}$$
$$= A \quad \text{if 0 not in B, } 1 \in B$$
$$= A^c \quad \text{if } 0 \in B \text{, 1 not in B}$$
$$= \Omega \quad \text{if } 0 \in B, 1 \in B$$

Then, $\sigma(I_A) = \{\emptyset, A, A^c, \Omega\}$ and I_A is Borel if and only if $A \in \Sigma$.

$\sigma(f)$ is much simpler than Σ.

• Note: We express the elements of Ω as numbers rather than abstract elements of sets.

Measurable Function – Properties

Theorems: Let (Ω, Σ) be a measurable space.

(i) f is a RV if and only if $f^{-1}(a, \infty) \in \Sigma$ for all $a \in R$.

(ii) If f and g are RVs, then so are fg and $af + bg$, where a and $b \in R$; also, f/g is a RV provided $g(\omega) \neq 0$ for any $\omega \in \Omega$.

(iii) If $f_1, f_2, \ldots$ are RVs, then so are $sup_n f_n$, $inf_n f_n$, $lim\ sup_n f_n$, and $lim\ inf_n f_n$. Furthermore, the set

$$A = \{\omega \in \Omega: \lim_{n\to\infty} f_n(\omega) \text{ exists}\}$$

is an event and the function

$$h(\omega) = \lim_{n\to\infty} f_n(\omega) \qquad \omega \in A$$
$$= f_1(\omega) \qquad \omega \in A^c$$

is a RV.

Measurable Function – Properties

Theorems: Let (Ω, Σ) be a measurable space.

(iv) (Closed under composition) Suppose that f is measurable from (Ω, Σ) to (Λ, G) and g is measurable from (Λ, G) to (Δ, H). Then, the composite function $g \circ f$ is measurable from (Ω, Σ) to (Δ, H).

(v) Let Ω be a Borel set in R^p. If f is a continuous function from Ω to R^p, then f is measurable.

Distribution

Definition

Let (Ω, Σ, μ) be a measure space and f be a measurable function from (Ω, Σ) to (Λ, G). The induced measure by f, denoted by $\mu \circ f^{-1}$, is a measure on G defined as

$$\mu \circ f^{-1}(B) = \mu(f \in B) = \mu(f^{-1}(B)), \qquad B \in G$$

If $\mu = P$ is a probability measure and X is a random variable or a random vector, then $P \circ X^{-1}$ is called the *distribution* (or the *law*) of X and is denoted by P_X.

• The cdf of P_X is also called the cdf (or joint cdf) of X and is denoted by F_X.

Probability Space – Definition and Axioms

Kolmogorov's axioms

Kolmogorov defined a list of axioms for a probability measure.

Let P: $E \rightarrow [0; 1]$ be our probability measure and E be some σ-algebra (events) generated by X.

Axiom 1: $P[A] \leq 1$ for all $A \in E$

Axiom 2. $P[X] = 1$

Axiom 3. $P[A_1 U A_2 U \ldots U A_n] = P[A_1] + P[A_2] + \ldots + P[A_n]$, where $\{A_1; A_2; \ldots; A_n\}$ are disjoint sets in E.

Probability Space – Properties of P

The three Kolmogorov's basic axioms imply the following results:

Theorem: $P[{A_i}^C] = 1 - P[A_i]$.

Theorem: $P[\Phi] = 0$

Theorem: $P[A_i] \in [0, 1]$.

Theorem: $P[B \cap A^C] = P[B] - P[A \cap B]$

Theorem: $P[A U B] = P[A] + P[B] - P[A \cap B]$

Theorem: A is in B $\Rightarrow P[A] \leq P[B]$

Theorem: $A = B$ $\Rightarrow P[A] = P[B]$

Theorem: $P[A] = \sum_{i=1} P[A \cap C_i]$, where $\{C_1; C_2; \ldots\}$ forms a partition of E.

Theorem *(Boole's Inequality, aka "Countable Subadditivity")*:

$$P[\cup_{i=1} A_i] \leq \sum_i^{\infty} P[A_i] \text{ for any set of sets } \{A_1; A_2; \ldots\}$$

Probability Space – (Ω, Σ, P)

Now, we have all the tools required to establish that (Ω, Σ, P) is a probability space.

Theorem:

Let Ω be the sample space of outcomes of an experiment, Σ be the σ-algebra of events generated from Ω, and P: Σ → [0, ∞) be a probability measure that assigns a nonnegative real number to each event in Σ. The space (Ω, Σ, P) satisfies the definition of a probability space.

Remark: The sample space is the list of all possible outcomes. *Events* are groupings of these outcomes. The σ-algebra Σ is the collection of all possible events. To each of these possible events, we assign some "size" using the probability measure P.

Probability Space

Example: Consider the weekly sign of stock returns of two unrelated firms: Positive (U: up) or negative (D: down).

The sample space is Ω = [{U,D}; {D,U}; {D,D} & {U,U}].

Possible events (*A*):

- Both firms have the same signed return: {U,U} & {D,D}.
- At least one firm has positive returns: {U,U}; {D,U} & {U,D}.
- The first firm is has positive returns: {U,U} & {U,D}

Collection of all possible events: Σ = [Φ, {U,U}, {U,D}, {D,U}, {D,D}, {UU, UD}, {UU, DU}, {UU, DD}, {DD, DU}, {DD, UD}, {DU, DD}, {UU, DU, UD}, {UD, DU, DD}, {UU, UD, DU, DD}]

The probability measure P assigns a number from 0 to 1 to each of those events in the sigma algebra.

Review – Random Variables – Example

Example (continuation): X takes Ω into χ & induces P_X from P.

Assuming equal probabilities for U & D, P[U] = P[D] = ½ :

Prob. of 0 Ups = $P_X[0]$ = P[{DD}] = ¼

Prob. of 1 Ups = $P_X[1]$ = P[{UD; DU}] = ½

Prob. of 2 Ups = $P_X[2]$ = P[{UU}] = ¼

Prob. of 0 or 1 Ups = $P_X[\{0; 1\}]$ = P[{DD; DU; UD}] = ¾

Prob. of 0 or 2 Ups = $P_X[\{0; 2\}]$ = P[{DD; UU}] = ½

Prob. of 1 or 2 Ups = $P_X[\{1; 2\}]$ = P[{UU; DU; DD}] = ¾

Prob. of 0, 1, or 2 Ups = $P_X[\{0; 1; 2\}]$ = P[{UU; DU; UD; DD}] = 1

Prob. of "nothing" = $P_X[\Phi]$ = P[Φ] = 0

The empty set is simply needed to complete the σ-algebra (a technical point). Its interpretation is not important since P[Φ] = 0.

27

Probability Space

Example (continuation):

Suppose we assign see D (a "bad" firm) 3/4 of the time. Then the following values for P would be appropriate:

{0; 9/16; 3/16; 3/16; 1/16; 3/4; 3/4; 5/8; 3/8; 1/4; 1/4; 15/16; 13/16; 13/16; 7/16; 1}

Probability Space

As long as the values of the probability measure are consistent with Kolmogorov's axioms and the consequences of those axioms, then we consider the probabilities to be mathematically acceptable, even if they are not reasonable for the given experiment.

Philosophical comment: Can the probability values assigned be considered reasonable as long as they're mathematically acceptable?

Random Variables Revisited

A random variable is a convenient way to express the elements of Ω as numbers rather than abstract elements of sets.

Definition: Measurable function

Let A_X; A_Y be nonempty families of subsets of X and Y, respectively. A function f: $X \rightarrow Y$ is $(A_X; A_Y)$-measurable if $f^{-1}(A) \in A_X$ for all $A \in A_Y$.

Definition: Random Variable

A *random variable* **X** is a measurable function from the probability space (Ω, Σ, P) into the probability space (χ, A_X, P_X), where χ in R is the range of **X** (which is a subset of the real line) A_X is a Borel field of X, and P_X is the probability measure on χ induced by **X**.

Specifically, **X**: $\Omega \rightarrow \chi$.

Random Variables Revisited

Example (continuation): Back to the weekly sign of stock returns of two unrelated firms: Positive (U: up) or negative (D: down).

Collection of all possible events:

Σ = [Φ, {U,U}, {U,D}, {D,U}, {D,D}, {UU, UD}, {UU, DU}, {UU, DD}, {DD, DU}, {DD, UD}, {DU, DD}, {UU, DU, UD}, {UD, DU, DD}, {UU, UD, DU, DD}]

Define RV: **X** = *"Number of Up cycles."* Recall, **X** takes Ω into χ,

χ = {0; 1; 2} and Σ_χ = {Φ; {0}; {1}; {2}; {0;1}; {0;2}; {1;2}; {0;1;2}}.

Then, **X**: $\Omega \rightarrow \chi$

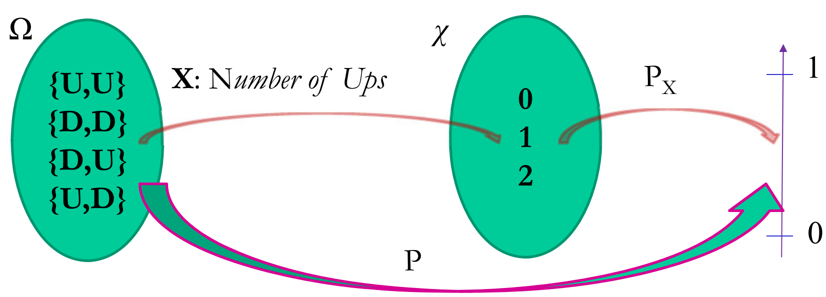
Then, we associate the elements in χ with a probability, P_X.

Random Variables Revisited

Example (continuation):

Then, **X**: $\Omega \rightarrow \chi$

Then, we associate the elements in χ with a probability, P_X.



In this example, χ = {0; 1; 2}

Σ_χ = {Φ; {0}; {1}; {2}; {0;1}; {0;2}; {1;2}; {0;1;2}}. 32