Econ 508B: Lecture 1

Probability and Distributions

Hongyi Liu

Washington University in St. Louis

August 2, 2018

Outline

Outline

1. Introduction
2. Set Theory
3. Fundamentals of Measure and Probability Set Function
4. Conditional Probability and Independence
5. Induced Measure, Random Variables and Distribution Functions

Probability v.s. Statistics

Probability v.s. Statistics

- **Probability example:** Flip a fair coin (equal probability of heads or tails) and toss it 10 times, what is the probability of occurring 8 or more heads? There is only one answer 7/128 and we center on how to compute it.
- **Statistics example:** You have a biased coin and intend to investigate whether it is fair when you toss it 100 times. Observe 55 heads, your objective is to draw a conclusion or inference from sample data.

Probability v.s. Statistics

Probability

As for the first example, the random process is fully given, e.g. probability of heads = 0.5. And the object is to find the probability of event 'occurring 8 or more heads'.

Probability v.s. Statistics

Probability

As for the first example, the random process is fully given, e.g. probability of heads = 0.5. And the object is to find the probability of event 'occurring 8 or more heads'.

Statistics

However, for the second case, the sample outcome is known (55 heads), but we need to illustrate the random process for this biased coin (probability of heads).

Suppose that we have such an experiment as follows

- the outcome of which cannot be predicted with certainty, but the experiment is of such a nature that a collection of every possible outcome can be described prior to its performance.
- If this experiment can be repeated under the same conditions, it is called a ***random experiment***, and the collection of every possible outcome is called the ***sample space***.

Suppose that we have such an experiment as follows

- the outcome of which cannot be predicted with certainty, but the experiment is of such a nature that a collection of every possible outcome can be described prior to its performance.
- If this experiment can be repeated under the same conditions, it is called a ***random experiment***, and the collection of every possible outcome is called the ***sample space***.

Example 1.1

Consider tossing of a coin twice, let the outcome tails be T and let the outcome heads be H. If we assume that the coin could be repeatedly tossed under the same conditions, therefore, the outcome is one of TT,TH,HT,HH and the sample space is the collection of these, e.g., $\{TT, TH, HT, HH\}$.

Let Ω denote a sample space, let ω denote an element of Ω, and let C represent a collection of elements of Ω.

- If the outcome is in C, we say that the ***event*** C has occurred.
- N repeated trials of the random experiment and the number f of times(the frequency) that the event C occurred throughout the N trials, the ratio of f/N is called the ***relative frequency*** of the event C in these N experiments.
- As N increases to a very large number, experience indicates that we associate with the event C a number, p, that is equal or approximately equal to that number about which the relative frequency seems to stabilize, and p is also referred to as the ***probability*** of the event C.

Outline

Set up

Ω abstractly representing the sample space of some experiments; ω, an individual element of Ω.

- $\mathcal{P}(\Omega) \equiv \{C : C \subset \Omega\}$: the power set of Ω. The power set of Ω is the set of all subsets of Ω; it includes the empty set and Ω itself.
- subsets of Ω mostly will be thought of as *events*, i.e., collections of simple events (points of Ω)
- collection of subsets will be written by calligraphic letters, e.g., $\mathcal{C}$
- the empty set or the null set is denoted by $\emptyset$.

Set operation

$\{C_k, k \in \mathbb{N}^+\}$ are collections of elements, we use the following notation,

- **Union:** The set of all elements that belong to one of the sets C_1 and C_2 is called the union of C_1 and C_2, which is indicated by $C_1 \cup C_2$. The union of several sets $C_1, C_2, C_3, ...$ is the set of all elements that are in one of the several sets, $C_1 \cup C_2 \cup C_3 \cup ...$ or by $C_1 \cup C_2 \cup ... \cup C_k$ if a finite number k of sets is involved.
- **Intersection:** The set of all elements that belong to each of the sets C_1 and C_2 is called the intersection of C_1 and C_2, which is indicated by $C_1 \cap C_2$. The intersection of several sets $C_1, C_2, C_3, ...$ is the set of all elements that are in each of the several sets, $C_1 \cap C_2 \cap C_3 \cap ...$ or by $C_1 \cap C_2 \cap ... \cap C_k$ if a finite k involved.
- **Complementation:** The set consists of all elements of Ω that are excluded from C is called the complement of C, denoted by C^c.

continued

Example 2.1 (DeMorgan's Laws)

$(C_1 \cap C_2)^c = C_1^c \cup C_2^c$, $(C_1 \cup C_2)^c = C_1^c \cap C_2^c$.

- **Monotone limits:** If $\{C_k\}$ is a monotone sequence of sets in $\mathcal{C}$ (a collection of subsets of Ω), the monotone limit $\lim_{k\to\infty} C_k$ is $\cup_{k=1}^{\infty} C_k$ when $\{C_k\}$ is non-decreasing and is $\cap_{k=1}^{\infty} C_k$ if $\{C_k\}$ is non-increasing.

Definition 2.1 (Closure)

Let $\mathcal{C}$ be a collection of subsets of Ω. $\mathcal{C}$ is closed under one of the above set operations if the set obtained by performing the set operation on sets in $\mathcal{C}$ yields a set in $\mathcal{C}$.

Example 2.2

Suppose $\Omega = \mathbb{R}$ and $\mathcal{C} = \{(a, b], -\infty < a \leqslant b < \infty\}$.

- $\mathcal{C}$ is not closed under finite unions since, e.g., $(1, 2] \cup (3, 4]$ is not an interval of the type defined by $\mathcal{C}$.
- However, $\mathcal{C}$ is closed under finite intersections since $(a, b] \cap (c, d] = (a \vee c, d \wedge b]$.

Example 2.3

Suppose $\Omega = \mathbb{R}$ and $\mathcal{C}$ consists of the open subset of $\mathbb{R}$.

- $\mathcal{C}$ is not closed under complementation since the complement of an open set is not open.

Why do we need closure?

A probability model has an event space and this is the class of subsets of sample space Ω to which we know how to assign probabilities.

- We cannot always assign probabilities to all of subsets. Therefore, we need to elicit a class of subsets that have assigned probabilities, which are referred to as **events**.
- obtaining complex events from simple events by set operation.
- The closure of certain set operation could guarantee that events outside the event space would not occur.

Outline

1. Introduction
2. Set Theory
3. Fundamentals of Measure and Probability Set Function
4. Conditional Probability and Independence
5. Induced Measure, Random Variables and Distribution Functions

Algebra(Field)

Definition 3.1 (Algebra)

An **algebra** is a collection $\mathcal{C}$ of subsets of $\mathcal{P}(\Omega)$ such that

(a) $\emptyset \in \mathcal{C}$ and $\Omega \in \mathcal{C}(\Omega$ is not empty).

(b) $C \in \mathcal{C}$ implies $C^c \in \mathcal{C}$, and

(c) $C_1, ..., C_k \in \mathcal{C}$ implies $\cup_{i=1}^{k} C_i$ (and $\cap_{i=1}^{k} C_i$) are also in $\mathcal{C}$.

Algebra(Field)

Definition 3.1 (Algebra)

An **algebra** is a collection $\mathcal{C}$ of subsets of $\mathcal{P}(\Omega)$ such that

(a) $\emptyset \in \mathcal{C}$ and $\Omega \in \mathcal{C}$(Ω is not empty).

(b) $C \in \mathcal{C}$ implies $C^c \in \mathcal{C}$, and

(c) $C_1, ..., C_k \in \mathcal{C}$ implies $\cup_{i=1}^{k} C_i$ (and $\cap_{i=1}^{k} C_i$) are also in $\mathcal{C}$.

Therefore, an algebra is a class of sets which contains sample space Ω and is closed under a finite number of set operations, namely we could establish a new set from a finite number of sets in the algebra $\mathcal{C}$ using a finite number of set operations (union, intersection and complement), then this new set is also in $\mathcal{C}$. However, this formulation is not enough for probability, since the collection of admissible subsets is usually defined to be closed under a **countable** number of set operations.

σ-Algebra

Definition 3.2 (σ-Algebra)

A class $\mathcal{C} \in \mathcal{P}(\Omega)$ is a σ-algebra if it is an algebra and it satisfies

(c*) whenever $C_1, C_2 ... \in \mathcal{C} \in \mathcal{C}$, then $\cup_{i=1}^{\infty} C_i$ (and $\cap_{i=1}^{\infty} C_i$) are also in $\mathcal{C}$.

A σ-algebra is a class of subsets of Ω that contains Ω and is closed under complementation, countable union and countable intersection.

Remark 3.1

- Uncountable unions and intersections are not allowed.
- $\cap_{i=1}^{\infty} C_i = (\cup_{i=1}^{\infty} C_i^c)^c$, hence, requiring $\cap_{i=1}^{\infty} C_i \in \mathcal{C}$ is redundant.

In probability theory, the event space is a σ-algebra.

Example 3.1

Let $\Omega = \{a, b, c, d\}$, then $\mathcal{C}_1 = \{\Omega, \emptyset, \{a\}\}$ and $\mathcal{C}_2 = \{\Omega, \emptyset, \{a\}, \{b, c, d\}\}$. where $\mathcal{C}_1$ is not an algebra, while $\mathcal{C}_2$ is both an algebra and σ-algebra.

Example 3.2

Ω is any non-empty set and $\mathcal{C}_3 = \mathcal{P}(\Omega) \equiv \{C : C \subset \Omega\}$ and $\mathcal{C}_4 = \{\Omega, \emptyset\}$. where $\mathcal{C}_3$ and $\mathcal{C}_4$ are both σ-algebras.

Definition 3.3

If $\mathcal{A}$ is a class of subsets of Ω, then the σ-algebra generated by $\mathcal{A}$, denoted by $\sigma\langle\mathcal{A}\rangle$, is defined as

$$\sigma\langle\mathcal{A}\rangle = \bigcap_{\mathcal{C} \in \mathcal{I}(\mathcal{A})} \mathcal{C}$$

where $\mathcal{I}(\mathcal{A}) \equiv \{\mathcal{C} : \mathcal{A} \subset \mathcal{C} \text{ and } \mathcal{C} \text{ is a } \sigma\text{-algebra on } \Omega\}$ is the collection of all σ-algebras containing the class $\mathcal{A}$.

Motivation

Example 3.3

Let $\Omega = \{a, b, c, d\}$, consider $\mathcal{C}_1 = \{\Omega, \emptyset, \{a\}\}, \mathcal{C}_2 = \{\Omega, \emptyset, \{a\}, \{b, c, d\}\}$, We have $\sigma\langle\mathcal{C}_1\rangle = \mathcal{C}_2$.

- In a sense, in probability model, we begin with a restricted class of subsets $\mathcal{C}$ to which we know how to assign probabilities.
- Then in principle, we know how to assign probabilities to $\sigma\langle\mathcal{C}\rangle$ via a *countable* collection of set operations, such manipulations may generate events outside $\mathcal{C}$, but not outside $\sigma\langle\mathcal{C}\rangle$.

Example 3.4

Let $\Omega = (0, 1]$, take $\mathcal{C} = \{(a, b], 0 \leqslant b \leqslant 1\}$ and $P((a, b]) = b - a$.

Borel sets and Borel σ-algebra

Definition 3.4 (Borel σ-algebra)

The **Borel σ-algebra** on a topological space $\mathbb{S}$ (e.g., on a metric space or an Euclidean space) is defined as the σ-algebra generated by the collection of open sets in $\mathbb{S}$, namely is the smallest σ-algebra containing all open sets of real numbers.

Definition 3.5 (Borel set)

A **Borel set** is any set in a topological space that can be formed from open sets (or, equivalently, from closed sets) through the operations of countable union, countable intersection, and relative complement.

Example 3.5

Let $\Omega = \mathbb{R}$, and take $\mathcal{C} = \{(a, b], -\infty \leqslant a \leqslant b < 1\}$ and define $\mathcal{B}(\mathbb{R}) = \sigma\langle\mathcal{C}\rangle$. Here $\mathcal{B}(\mathbb{R})$ is called the borel subsets of $\mathbb{R}$.

What do borel sets look like? Examples

Definition 3.6

$\{G\}$ denotes the open sets in R^n and $\{F\}$ denotes the closed sets. Hence, H is of type G_δ if

$$H = \bigcap_k G_k, G_k \text{ open}$$

and H is of type F_σ if

$$H = \bigcup_k F_k, F_k \text{ closed}$$

- Every G_δ set is a Borel set. Since the complement of a G_δ set is an F_σ, then every F_σ set is also a Borel set.
- Every interval of the form $[a, b)$ is both a G_δ set and an F_σ set, therefore, it is a Borel set.

Retrospect

Recall the last example, in fact, there are many equivalent ways of generating the Borel sets as follows:

$$
\begin{aligned}
\mathcal{B}(\mathbb{R}) &= \sigma\langle (a,b), -\infty \leq b \leq \infty \rangle \\
&= \sigma\langle [a,b), -\infty < a \leq b \leq \infty \rangle \\
&= \sigma\langle [a,b], -\infty < a \leq b < \infty \rangle \\
&= \sigma\langle (-\infty, x], x \in \mathbb{R} \rangle \\
&= \sigma\langle open\ subsets\ of\ \mathbb{R} \rangle
\end{aligned}
$$

Therefore, the Borel sets with any kind of interval: open, closed, semi-open, finite, semi-infinite, etc.

Terminology Review

- By a *set function* we mean a function which assigns an extended real number to class of subsets of a set Ω.
- By a *measurable space* we mean $(\Omega, \mathcal{C})$ consisting of a set Ω and a σ-algebra $\mathcal{C}$ of subsets of Ω. A subset of C of Ω is called measurable if $C \in \mathcal{C}$. In statistics and probability terminology, events are measurable set.
- $(\Omega, \mathcal{C}, \mu)$ is referred to as a *measure space* when μ is a measure on the measurable space $(\Omega, \mathcal{C})$.

Measure

Let Ω be a nonempty set and $\mathcal{C}$ be a σ-algebra on Ω. Then, a set function μ on $\mathcal{C}$ is called a **measure** if

- $\mu(C) \in [0, +\infty]$ for all $C \in \mathcal{C}$,
- $\mu(\emptyset) = 0$,
- for any countable $\{C_n\}$, pairwise disjoint collection of sets in $\mathcal{C}$.

$$\mu(\bigcup_{n\geq 1} C_n) = \sum_{n=1}^{\infty} \mu(C_n)$$

Remark 3.2

The third property is referred to as *countable additivity*, which is equivalent to conditions: **finite additivity** and **monotone continuity from below**.

continued

Proposition 3.1

Let Ω be a nonempty set and $\mathcal{C}$ be an algebra of subsets of Ω and μ be a set function on $\mathcal{C}$ with values in $[0, \infty]$ and with $\mu(\emptyset) = 0$. Then μ is a measure if.f. μ satisfies

- **(finite additivity:)** for all $C_1, C_2 \in \mathcal{C}$ with $C_1 \cap C_2 = \emptyset$, $\mu(C_1 \cup C_2) = \mu(C_1) + \mu(C_2)$,
- **(monotone continuity from below:)** for any collection $\{C_n\}_{n \geq 1}$ of sets in $\mathcal{C}$ such that $C_n \subset C_{n+1}$ for all $n \geq 1$ and $\bigcup_{n \geq 1} C_n \in \mathcal{C}$, $\mu(\bigcup_{n \geq 1} C_n) = \lim_{n \to \infty} \mu(C_n)$

Probability measures and finite measures

Definition 3.7

A measure μ is called finite or infinite according to whether $\mu(\Omega) < \infty$ or $\mu(\Omega) = \infty$.

- **Probability measure:** $\mu(\Omega) = 1 (P(\Omega) = 1)$.

Definition 3.8

A measure μ on a σ-algebra $\mathcal{C}$ is called ***σ-finite*** if there exists a countable collection of sets $C_1, C_2, ... \in \mathcal{C}$, not necessarily disjoint, such that

- $\bigcup_{n \geq 1} C_n = \Omega$, and
- $\mu(C_n) < \infty$ for all $n \geq 1$.

Example 3.6 (Counting measure)

Let Ω be a nonempty set and $\mathcal{C}$ be the set of all subsets of Ω. Define

$$\mu(C) = |C|, C \in \mathcal{C},$$

where $|C|$ denotes the number of elements in C.

- It is easy to check that here μ satisfies with three prerequisites for a measure.
- Note that μ is finite if.f. Ω is finite and it is σ-finite if Ω is countably infinite.

Proposition 3.2

Let μ be a measure on an algebra $\mathcal{C}$ and let $C, D, C_1, ..., C_k \in \mathcal{C}$, where $1 \leq k < \infty$. Then

(1) (**Monotonicity:**) $\mu(C) \leq \mu(D)$ if $C \subset D$,

(2) (**Finite subadditivity:**) $\mu(C_1 \cup ... \cup C_k) \leq \mu(C_1) + ... + \mu(C_k)$,

(3) (**Inclusion-exclusion formula**) If $\mu(C_i) < \infty$ for all $i = 1, ..., k$

$$\mu(C_1 \cup ... \cup C_k) = \sum_{i=1}^{k} \mu(C_i) - \sum_{1 \leq i < j < k} \mu(C_i \cap C_j) + ... + (-1)^{k-1} \mu(C_1 \cap ... \cap C_k)$$

- *Question:* if μ is a measure on $\mathcal{C}$ and $\{C_n\}_{n \geq 1}$ is a collection of decreasing sets in $\mathcal{C}$ with $C \equiv \bigcap_{n \geq 1} C_n$ also in $\mathcal{C}$, does the relation $\mu(C) = \lim_{n \to \infty} \mu(C_n)$ hold?
- Answer: positive if imposed by the assumption $\mu(C_{n_0}) < \infty$ for some $n_0 \in \mathbb{N}$.

Proposition 3.3

(1) (**Monotone continuity from above :**) Let $\{C_n\}_{n\geq 1}$ be a sequence of sets in $\mathcal{C}$ such that $C_{n+1} \subset C_n$ for all $n \geq 1$ and $C \equiv \bigcap_{n\geq 1} C_n \in \mathcal{C}$. Also, let $\mu(C_{n_0}) < \infty$ for some $n_0 \in \mathbb{N}$. Then, $\mu(C) = \lim_{n\to\infty} \mu(C_n)$.

(2) (**Countable subadditivity:**) If $\{C_n\}_{n\geq 1}$ is a sequence of sets in $\mathcal{C}$ such that $\bigcup_{n\geq 1} C_n \in \mathcal{C}$, then

$$\mu(\bigcup_{n\geq 1} C_n) \leq \sum_{n=1}^{\infty} \mu(C_n)$$

Application in probability measure

When $\mu(\Omega) = P(\Omega) = 1$, all of the above properties could be applied into probability measure. In fact, there are several particular characteristics in probability measure.

- For each event $C \in \mathcal{C}, P(C) = 1 - P(C^c)$,
- $P(\emptyset) = 0$,
- *Boole's inequality:*

$$1 \geq P(C_1 \cup C_2) = P(C_1) + P(C_2) - P(C_1 \cap C_2)$$

Outline

Conditional probability

- In some random experiments, we are interested only in those outcomes that are elements of a subset C_1 of the sample space Ω.

Conditional probability

- In some random experiments, we are interested only in those outcomes that are elements of a subset C_1 of the sample space Ω.

Q:How to define the "new" sample space C_1 *and the conditional probability?*

Conditional probability

- In some random experiments, we are interested only in those outcomes that are elements of a subset C_1 of the sample space Ω.

Q:How to define the "new" sample space C_1 and the conditional probability?

Definition 4.1

Let $P(\cdot)$ be defined on the sample space Ω and let $C_1, C_2 \in \Omega$ such that $P(C_1) > 0$.

$$P(C_1|C_1) = 1 \text{ and } P(C_2|C_1) = P(C_1 \cap C_2|C_1).$$

And from a relative frequency point of view, we should have

$$P(C_1 \cap C_2|C_1) = \frac{P(C_1 \cap C_2|C_1)}{P(C_1|C_1)} = \frac{P(C_1 \cap C_2)}{P(C_1)}$$

Moreover,

- $P(C_2|C_1) \geq 0$,
- $P(C_2 \cup C_3...|C_1) = P(C_2|C_1) + P(C_3|C_1) + ...$, provided that $C_2, C_3, ...$ are mutually disjoint sets.
- *multiplication rule:* $P(C_1 \cap C_2) = P(C_1)P(C_2|C_1)$.

Consider k mutually exclusive events $C_1, C_2, ..., C_k$ such that $P(C_i) > 0, i = 1, 2, ..., k$. Suppose that these events form a partition of $\mathcal{C}$. Here the events $C_1, C_2, ..., C_k$ do not need to be equally likely. Let C be another event. Thus C occurs with one and only one of the events $C_1, C_2, ...C_k$; that is,

$$\begin{aligned} C &= C \cap (C_1 \cup C_2 \cup ...C_k) \\ &= (C \cap C_1) \cup (C \cap C_2) \cup ... \cup (C \cap C_k). \end{aligned}$$

Law of total probability & Bayes' theorem

Since $C \cap C_i, i = 1, 2, ..., k$, are mutually exclusive, then

$$P(C) = P(C \cap C_1) + P(C \cap C_2) + ... + P(C \cap C_k).$$

Furthermore, $P(C_1 \cap C_2) = P(C_1)P(C_2|C_1), i = 1, 2, ..., k$, so

$$\begin{aligned} P(C) &= P(C_1)P(C|C_1) + P(C_2)P(C|C_2) + ... + P(C_k)P(C|C_k) \\ &= \sum_{i=1}^{k} P(C_i)P(C|C_i) \qquad \textit{(law of total probability)} \end{aligned}$$

Suppose that $P(C) > 0$, from the definition of conditional probability

$$P(C_j|C) = \frac{P(C \cap C_j)}{P(C)} = \frac{P(C_j)P(C|C_j)}{\sum_{i=1}^{k} P(C_i)P(C|C_i)} \qquad \textit{(Bayes' theorem)}$$

Independence

Definition 4.2

Let C_1 and C_2 be two events in $\mathcal{C}$. C_1 and C_2 are independent if

$$P(C_1 \cap C_2) = P(C_1)P(C_2)$$

Moreover,

$$P(C_2|C_1) = \frac{P(C_1 \cap C_2)}{P(C_1)} = P(C_2)$$
$$P(C_1|C_2) = \frac{P(C_1 \cap C_2)}{P(C_2)} = P(C_1)$$

Outline

Kolmogorov's Probability Model

Impose that $\mathcal{C}$ and P satisfy the following conditions:

- $C \in \mathcal{C} \Rightarrow C^c \in \mathcal{C}$,
- $C_1, C_2 \in \mathcal{C} \Rightarrow C_1 \cup C_2 \in \mathcal{C}$,
- for all $C \in \mathcal{C}, 0 \leqslant P(C) \leq 1, P(\emptyset) = 0$ and $P(\Omega) = 1$,
- $C_1, C_2 \in \mathcal{C}, C_1 \cap C_2 = \emptyset \Rightarrow P(C_1 \cup C_2) = P(C_1) + P(C_2)$,
- $C_n \in \mathcal{C}, C_n \subset C_{n+1}$ for all $n = 1, 2, \ldots \Rightarrow \bigcup_{n \geq 1} C_n \in \mathcal{C}$ and $P(C_n) \uparrow P(\bigcup_{n \geq 1} C_n)$.

Remark 5.1

The above all conditions imply $(\Omega, \mathcal{C}, P)$ is a **measure space**, i.e. $\mathcal{C}$ is σ-algebra, P is a measure on $\mathcal{C}$ with $P(\Omega) = 1$. Thus $(\Omega, \mathcal{C}, P)$ is a *probability space*.

How to measure over sample spaces

Example 5.1 (Finite sample spaces)

Let $\Omega = \{\omega_1, ..., \omega_k\}, 1 \leq k < \infty, \mathcal{C} \equiv \mathcal{P}(\Omega)$ and $P(C) = \sum_{i=1}^{k} p_i \mathbb{1}_C(\omega_i)$, where $\{p_i\}_{i=1}^{k}$ are such that $p_i \geq 0$ and $\sum_{i=1}^{k} p_i = 1$. This is a probability model for random experiments with finitely many possible outcomes.

How to measure over sample spaces

Example 5.1 (Finite sample spaces)

Let $\Omega = \{\omega_1, ..., \omega_k\}, 1 \leq k < \infty, \mathcal{C} \equiv \mathcal{P}(\Omega)$ and $P(C) = \sum_{i=1}^{k} p_i \mathbb{1}_C(\omega_i)$, where $\{p_i\}_{i=1}^{k}$ are such that $p_i \geq 0$ and $\sum_{i=1}^{k} p_i = 1$. This is a probability model for random experiments with finitely many possible outcomes.

Example 5.2 (Countably infinite sample spaces)

Let $\Omega = \{\omega_1, \omega_2, ...\}$ be a countable set, $\mathcal{C} \equiv \mathcal{P}(\Omega)$ and $P(C) = \sum_{i=1}^{\infty} p_i \mathbb{1}_C(\omega_i)$, where $\{p_i\}_{i=1}^{\infty}$ are such that $p_i \geq 0$ and $\sum_{i=1}^{\infty} p_i = 1$. Moreover, $(\Omega, \mathcal{C}, P)$ is a *probability space.*

- The experiment of tossing a coin until 'head' is observed leads to such a probability space.

Uncountable sample spaces

Example 5.3 (Random variable)

Let $\Omega = \mathbb{R}, \mathcal{C} = \mathcal{B}(\mathbb{R}), P = \mu_F$, the Lebesgue-Stieltjes measure corresponding to a cdf F, i.e. corresponding to a function $F : \mathbb{R} \to \mathbb{R}$ that is non-decreasing, right-continuous and satisfies $F(-\infty) = 0, F(+\infty) = 1$. This is a model for a single random variable X.

Example 5.4 (Random vectors)

Let $\Omega = \mathbb{R}^k, \mathcal{C} = \mathcal{B}(\mathbb{R}^k), P = \mu_F$, the Lebesgue-Stieltjes measure corresponding to a multidimensional cdf F on $\mathbb{R}^k$, where $k \in \mathbb{N}$. This is a model for a random vector $(X_1, X_2, ..., X_k)$.

It also suffices to show the uncountable samples spaces for random sequences.

Induced Measure

Proposition 5.1

Let $(\Omega_i, \mathcal{C}_i), i = 1, 2$ be measurable spaces and let $T : \Omega_1 \to \Omega_2$ be a $\langle \mathcal{C}_1, \mathcal{C}_2 \rangle$-measurable mapping from Ω_1 to Ω_2. Then, for any measure μ on $(\Omega_1, \mathcal{C}_1)$, the set function μT^{-1}, defined by

$$\mu T^{-1}(C) \equiv \mu(T^{-1}(C)), C \in \mathcal{C}_2$$

is measured on $\mathcal{C}_2$

Definition 5.1

The measure μT^{-1} is called the **measure induced** by T on $\mathcal{C}_2$.

Random Variable & Cumulative Distribution Function

Definition 5.2

Let $(\Omega, \mathcal{C}, P)$ be a probability space and $X : \Omega \to \mathbb{R}$ be $\langle \mathcal{C}, \mathcal{B}(\mathbb{R}) \rangle$-measurable, i.e. $X^{-1}(C) \in \mathcal{C}$ for all $C \in \mathcal{B}(\mathbb{R})$. Then X is called a **random variable** on $(\Omega, \mathcal{C}, P)$.

- Recall that $X : \Omega \to \mathbb{R}$ is $\langle \mathcal{C}, \mathcal{B}(\mathbb{R}) \rangle$-measurable iff for all $x \in \mathbb{R}$, $\{\omega : X(\omega) \leq x\} \in \mathcal{C}$.

Definition 5.3

Let X be a random variable on $(\Omega, \mathcal{C}, P)$. Let

$$F_X(x) \equiv P(\{\omega : X(\omega) \leq x\}) \equiv P_X((-\infty, x]), x \in \mathbb{R}.$$

Then $F_X(\cdot)$ is called the **cumulative distribution function** of X.

Proposition 5.2

Let F be the cdf of a random variable X.

- For $x_1 < x_2, F(x_1) \leq F(x_2)$ (i.e., F is nondecreasing on $\mathbb{R}$).
- For x in $\mathbb{R}$,$F(x) = \lim_{y \downarrow x} F(y)$ (i.e., F is right continuous on $\mathbb{R}$).
- $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{n \to +\infty} F(x) = 1$.

Definition 5.4

Let X be a random variable on $(\Omega, \mathcal{C}, P)$. Let

$$P_X(C) \equiv P(X^{-1}(C)) \ for \ all \ C \in \mathcal{B}(\mathbb{R})$$

Then the probability measure P_X is called the **probability distribution** of X.

Definition 5.5

Let $(\Omega, \mathcal{C}, P)$ be a probability space, $k \in \mathbb{N}$ and $X : \Omega \to \mathbb{R}^k$ be $\langle \mathcal{C}, \mathcal{B}(\mathbb{R}^k) \rangle$-measurable, i.e. $X^{-1}(C) \in \mathcal{C}$ for all $C \in \mathcal{B}(\mathbb{R}^k)$. Then X is called a k-dimensional **random vector** on $(\Omega, \mathcal{C}, P)$.

Definition 5.6

Let X be a k-dimensional random vector on $(\Omega, \mathcal{C}, P)$ for some $k \in \mathbb{N}$.

$$F_X(x) \equiv P(\{\omega : X_1(\omega) \leq x_1, X_2(\omega) \leq x_2, ..., X_k(\omega) \leq x_k\})$$

for $x = (x_1, ..., x_k) \in \mathbb{R}^k$. Then $F_X(\cdot)$ is referred to as the **joint cumulative distribution function** of the random vector X.

Definition 5.7

Let X be a k-dimensional **random vector** on $(\Omega, \mathcal{C}, P)$ for some $k \in \mathbb{N}$.

$$P_X(C) = P(X^{-1}(C)) \ for \ all \ C \in \mathcal{B}(\mathbb{R}^k).$$

The probability measure P_X is called the **joint probability distribution** of X.