Random experiment: uncertainty in outcomes
Ω: sample space: a set containing all possible outcomes

**Definition 1.1**

A collection of subsets of \( \Omega \), \( \mathcal{F} \), is a \( \sigma \)-field (or \( \sigma \)-algebra) if

(i) The empty set \( \emptyset \in \mathcal{F} \);
(ii) If \( A \in \mathcal{F} \), then the complement \( A^c \in \mathcal{F} \);
(iii) If \( A_i \in \mathcal{F}, i = 1, 2, ..., \) then their union \( \bigcup A_i \in \mathcal{F} \).

\((\Omega, \mathcal{F})\) is a measurable space if \( \mathcal{F} \) is a \( \sigma \)-field on \( \Omega \)

Two trivial examples: \( \mathcal{F} = \{\emptyset, \Omega\} \) and \( \mathcal{F} \) = all subsets of \( \Omega \) (power set)
A nontrivial example: \( \mathcal{F} = \{\emptyset, A, A^c, \Omega\} \), where \( A \subset \Omega \)

\( \mathcal{C} \) = a collection of subsets of interest (may not be a \( \sigma \)-field)

\( \sigma(\mathcal{C}) \) = the smallest \( \sigma \)-field containing \( \mathcal{C} \) (the \( \sigma \)-field generated by \( \mathcal{C} \))

\( \sigma(\mathcal{C}) = \mathcal{C} \) if \( \mathcal{C} \) itself is a \( \sigma \)-field

\( \sigma(\{A\}) = \sigma(\{A, A^c\}) = \sigma(\{A, \Omega\}) = \sigma(\{A, \emptyset\}) = \{\emptyset, A, A^c, \Omega\} \)
Borel $\sigma$-field

$\mathbb{R}^k$: the $k$-dimensional Euclidean space ($\mathbb{R}^1 = \mathbb{R}$ is the real line)

$\mathcal{O} =$ all open sets, $\mathcal{C} =$ all closed sets

$\mathcal{B}^k = \sigma(\mathcal{O}) = \sigma(\mathcal{C})$: the Borel $\sigma$-field on $\mathbb{R}^k$

$C \in \mathcal{B}^k$, $\mathcal{B}_C = \{ C \cap B : B \in \mathcal{B}^k \}$ is the Borel $\sigma$-field on $C$

Definition 1.2.

Let $(\Omega, \mathcal{F})$ be a measurable space.

A set function $\nu$ defined on $\mathcal{F}$ is a measure if

(i) $0 \leq \nu(A) \leq \infty$ for any $A \in \mathcal{F}$;

(ii) $\nu(\emptyset) = 0$;

(iii) If $A_i \in \mathcal{F}$, $i = 1, 2, \ldots$, and $A_i$'s are disjoint, i.e., $A_i \cap A_j = \emptyset$ for any $i \neq j$, then

$$\nu \left( \bigcup_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} \nu(A_i).$$

$(\Omega, \mathcal{F}, \nu)$ is a measure space if $\nu$ is a measure on $\mathcal{F}$ in $(\Omega, \mathcal{F})$.

If $\nu(\Omega) = 1$, then $\nu$ is a probability measure.

We usually use $P$ instead of $\nu$; i.e., $(\Omega, \mathcal{F}, P)$ is a probability space.
Conventions

For any $x \in \mathbb{R}$, $\infty + x = \infty$, $x \infty = \infty$ if $x > 0$, $x \infty = -\infty$ if $x < 0$

$0 \infty = 0$, $\infty + \infty = \infty$, $\infty^a = \infty$ for any $a > 0$;

$\infty - \infty$ or $\infty / \infty$ is not defined

Important examples of measures

Let $x \in \Omega$ be a fixed point and

$$\delta_x(A) = \begin{cases} 
    c & x \in A \\
    0 & x \notin A.
\end{cases}$$

This is called a point mass at $x$

Let $\mathcal{F}$ = all subsets of $\Omega$ and $\nu(A) =$ the number of elements in $A \in \mathcal{F}$ ($\nu(A) = \infty$ if $A$ contains infinitely many elements). Then $\nu$ is a measure on $\mathcal{F}$ and is called the counting measure.

There is a unique measure $m$ on $(\mathbb{R}, \mathcal{B})$ that satisfies $m([a, b]) = b - a$ for every finite interval $[a, b]$, $-\infty < a \leq b < \infty$. This is called the Lebesgue measure.

If we restrict $m$ to the measurable space $([0, 1], \mathcal{B}_{[0,1]})$, then $m$ is a probability measure (uniform distribution).
Proposition 1.1 (Properties of measures)

Let \((\Omega, \mathcal{F}, \nu)\) be a measure space.

1. (Monotonicity). If \(A \subset B\), then \(\nu(A) \leq \nu(B)\).

2. (Subadditivity). For any sequence \(A_1, A_2, \ldots\),

\[
\nu \left( \bigcup_{i=1}^{\infty} A_i \right) \leq \sum_{i=1}^{\infty} \nu(A_i).
\]

3. (Continuity). If \(A_1 \subset A_2 \subset A_3 \subset \cdots\) (or \(A_1 \supset A_2 \supset A_3 \supset \cdots\) and \(\nu(A_1) < \infty\)), then

\[
\nu \left( \lim_{n \to \infty} A_n \right) = \lim_{n \to \infty} \nu(A_n),
\]

where

\[
\lim_{n \to \infty} A_n = \bigcup_{i=1}^{\infty} A_i \quad \left( \text{or} = \bigcap_{i=1}^{\infty} A_i \right).
\]

Let \(P\) be a probability measure on \((\mathbb{R}, \mathcal{B})\).

The cumulative distribution function (c.d.f.) of \(P\) is defined to be

\[
F(x) = P((\neg \infty, x]) , \quad x \in \mathbb{R}
\]
Proposition 1.2 (Properties of c.d.f.'s)

(i) Let $F$ be a c.d.f. on $\mathbb{R}$.
   
   (a) $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$;

   (b) $F(\infty) = \lim_{x \to \infty} F(x) = 1$;

   (c) $F$ is nondecreasing, i.e., $F(x) \leq F(y)$ if $x \leq y$;

   (d) $F$ is right continuous, i.e., $\lim_{y \to x, y > x} F(y) = F(x)$.

(ii) Suppose a real-valued function $F$ on $\mathbb{R}$ satisfies (a)-(d) in part (i). Then $F$ is the c.d.f. of a unique probability measure on $(\mathbb{R}, \mathcal{B})$.

Product space

\[ \mathcal{I} = \{1, \ldots, k\}, \text{ } k \text{ is finite or } \infty \]
\[ \Gamma_i, i \in \mathcal{I}, \text{ are some sets} \]
\[ \prod_{i \in \mathcal{I}} \Gamma_i = \Gamma_1 \times \cdots \times \Gamma_k = \{(a_1, \ldots, a_k) : a_i \in \Gamma_i, i \in \mathcal{I}\} \]
\[ \mathbb{R} \times \mathbb{R} = \mathbb{R}^2, \mathbb{R} \times \mathbb{R} \times \mathbb{R} = \mathbb{R}^3 \]

Let $(\Omega_i, \mathcal{F}_i), i \in \mathcal{I}$, be measurable spaces

$\prod_{i \in \mathcal{I}} \mathcal{F}_i$ is not necessarily a $\sigma$-field

$\sigma(\prod_{i \in \mathcal{I}} \mathcal{F}_i)$ is called the product $\sigma$-field on the product space $\prod_{i \in \mathcal{I}} \Omega_i$

$(\prod_{i \in \mathcal{I}} \Omega_i, \sigma(\prod_{i \in \mathcal{I}} \mathcal{F}_i))$ is denoted by $\prod_{i \in \mathcal{I}} (\Omega_i, \mathcal{F}_i)$

Example: $\prod_{i=1,\ldots,k} (\mathbb{R}, \mathcal{B}) = (\mathbb{R}^k, \mathcal{B}^k)$
Product measure

Consider a rectangle $[a_1, b_1] \times [a_2, b_2] \subset \mathbb{R}^2$. The usual area of $[a_1, b_1] \times [a_2, b_2]$ is

$$(b_1 - a_1)(b_2 - a_2) = m([a_1, b_1])m([a_2, b_2])$$

Is $m([a_1, b_1])m([a_2, b_2])$ the same as the value of a measure defined on the product $\sigma$-field?

$\sigma$-finite

A measure $\nu$ on $(\Omega, \mathcal{F})$ is said to be $\sigma$-finite iff there exists a sequence $\{A_1, A_2, \ldots\}$ such that $\bigcup A_i = \Omega$ and $\nu(A_i) < \infty$ for all $i$.

Any finite measure (such as a probability measure) is clearly $\sigma$-finite. The Lebesgue measure on $\mathbb{R}$ is $\sigma$-finite, since $\mathbb{R} = \bigcup A_n$ with $A_n = (-n, n)$, $n = 1, 2, \ldots$.

The counting measure in is $\sigma$-finite if and only if $\Omega$ is countable. The measure $\nu(A) = \infty$ unless $A = \emptyset$ is not $\sigma$-finite.
Proposition 1.3 (Product measure theorem)

Let $(\Omega_i, \mathcal{F}_i, \nu_i), i = 1, \ldots, k$, be measure spaces with $\sigma$-finite measures. There exists a unique $\sigma$-finite measure on $\sigma$-field $\sigma(\mathcal{F}_1 \times \cdots \times \mathcal{F}_k)$, called the product measure and denoted by $\nu_1 \times \cdots \times \nu_k$, such that

$$\nu_1 \times \cdots \times \nu_k(A_1 \times \cdots \times A_k) = \nu_1(A_1) \cdots \nu_k(A_k)$$

for all $A_i \in \mathcal{F}_i, i = 1, \ldots, k$.

Joint and marginal c.d.f.'s

The joint c.d.f. of a probability measure on $(\mathbb{R}^k, \mathcal{B}^k)$ is defined by

$$F(x_1, \ldots, x_k) = P((-\infty, x_1] \times \cdots \times (-\infty, x_k]), \quad x_i \in \mathbb{R}$$

and the $i$th marginal c.d.f. is defined by

$$F_i(x) = \lim_{x_j \to \infty, j = 1, \ldots, i-1, i+1, \ldots, k} F(x_1, \ldots, x_i-1, x, x_i+1, \ldots, x_k)$$

There is a 1-1 correspondence between probability and c.d.f. on $\mathbb{R}^k$. The product measure corresponds to

$$F(x_1, \ldots, x_k) = F_1(x_1) \cdots F_k(x_k), \quad (x_1, \ldots, x_k) \in \mathbb{R}^k,$$

where $F_i$ is the c.d.f. of the $i$th probability measure.
Measurable function

$f$: a function from $\Omega$ to $\Lambda$ (often $\Lambda = \mathbb{R}^k$)

Inverse image of $B \subset \Lambda$ under $f$:

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

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

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

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

Let $\mathcal{C}$ be a collection of subsets of $\Lambda$.

Define $f^{-1}(\mathcal{C}) = \{ f^{-1}(C) : C \in \mathcal{C} \}$

Definition 1.3

Let $(\Omega, \mathcal{F})$ and $(\Lambda, \mathcal{G})$ be measurable spaces.

Let $f$ be a function from $\Omega$ to $\Lambda$.

$f$ is called a measurable function from $(\Omega, \mathcal{F})$ to $(\Lambda, \mathcal{G})$ iff $f^{-1}(\mathcal{G}) \subset \mathcal{F}$. 
Remarks

- $f$ is measurable from $(\Omega, \mathcal{F})$ to $(\Lambda, \mathcal{G})$ iff for any $B \in \mathcal{G}$, $f^{-1}(B) = \{\omega : f(\omega) \in B\} \in \mathcal{F}$; we don’t care about whether $\{f(\omega) : \omega \in A\}$ is in $\mathcal{G}$ or not, $A \in \mathcal{F}$.

- If $\mathcal{F} = \text{all subsets of } \Omega$, then any function $f$ is measurable.

- If $f$ is measurable from $(\Omega, \mathcal{F})$ to $(\Lambda, \mathcal{G})$, then $f^{-1}(\mathcal{G})$ is a sub-$\sigma$-field of $\mathcal{F}$ and is called the $\sigma$-field generated by $f$ and denoted by $\sigma(f)$.

- $\sigma(f)$ may be much simpler than $\mathcal{F}$.

- A measurable $f$ from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B})$ is called a Borel function.

- $f$ is Borel if and only if $f^{-1}(a, \infty) \in \mathcal{F}$ for all $a \in \mathbb{R}$.

- There are a lot of Borel functions (Proposition 1.4)

- In probability and statistics, a Borel function is also called a random variable. (A random variable = a variable that is random?)

- A random vector $(X_1, \ldots, X_n)$ is a function measurable from $(\Omega, \mathcal{F})$ to $(\mathbb{R}^n, \mathcal{B}^n)$.

- $(X_1, \ldots, X_n)$ is a random vector iff each $X_i$ is a random variable.
Indicator and simple functions

The indicator function for $A \subset \Omega$ is:

$$I_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A. \end{cases}$$

For any $B \subset \mathbb{R}$,

$$I_A^{-1}(B) = \begin{cases} \emptyset & 0 \notin B, 1 \notin B \\ A & 0 \notin B, 1 \in B \\ A^c & 0 \in B, 1 \notin B \\ \Omega & 0 \in B, 1 \in B. \end{cases}$$

Then, $\sigma(I_A) = \{\emptyset, A, A^c, \Omega\}$ and $I_A$ is Borel iff $A \in \mathcal{F}$

Note that $\sigma(I_A)$ is much simpler than $\mathcal{F}$.

Let $A_1, \ldots, A_k$ be measurable sets on $\Omega$ and $a_1, \ldots, a_k$ be real numbers. A simple function is

$$\varphi(\omega) = \sum_{i=1}^{k} a_i I_{A_i}(\omega)$$

A simple function is nonnegative iff $a_i \geq 0$ for all $i$.

Any nonnegative Borel function can be the limit of a sequence of nonnegative simple functions.
Let $A_1, \ldots, A_k$ be a partition of $\Omega$, i.e., $A_i$’s are disjoint and $A_1 \cup \cdots \cup A_k = \Omega$.
Then the simple function $\varphi$ with distinct $a_i$’s exactly characterizes this partition and $\sigma(\varphi) = \sigma(\{A_1, \ldots, A_k\})$.

### Distribution (law)

Let $(\Omega, \mathcal{F}, \nu)$ be a measure space and $f$ be a measurable function from $(\Omega, \mathcal{F})$ to $(\Lambda, \mathcal{G})$.
The induced measure by $f$, denoted by $\nu \circ f^{-1}$, is a measure on $\mathcal{G}$ defined as

$$\nu \circ f^{-1}(B) = \nu(f \in B) = \nu\left(f^{-1}(B)\right), \quad B \in \mathcal{G}$$

If $\nu = P$ is a probability measure and $X$ is a random variable or a random vector, then $P \circ X^{-1}$ is called the law or the distribution of $X$ and is denoted by $P_X$.
The c.d.f. of $P_X$ is also called the c.d.f. or joint c.d.f. of $X$ and is denoted by $F_X$. 