# Topic 5: Measure, Counting, Independence [1]

## Random experiments, outcomes and Events

A **random experiment** is a process that, when repeated under controlled conditions, does not always produce the same outcome. Before performing the experiment, we cannot determine which of the possible outcomes will occur.

Tossing a coin or rolling a die are classical examples of random experiments. However, the daily change in a stock market prices index or the hourly wage of a randomly selected individual are also examples of random experiments.

Although the result of a random experiment is unknown in advance, we can define the set of all possible outcomes it may produce.

**Definition 1** (Sample Space)

The set $S$ of all possible outcomes of a particular experiment is called the **sample space** of the experiment.

Sample spaces can be classified as countable or uncountable. A sample space is countable if its elements can be placed in one-to-one correspondence with a subset of the integers. This classification is important because it influences how probabilities are assigned.

We often consider collections of possible outcomes of a random experiment.

**Definition 2** (Event)

An **event** $A$ is any collection of possible outcomes of an experiment, that is, any subset of the sample space $S$.

An event $A$ is said to *occur* if the outcome of the experiment is in the set $A$.

[1] Instructors: Camilo Abbate and Sofia Olguin. This note was prepared for the 2025 UCSB Math Camp for Ph.D. students in economics. It incorporates materials from previous instructors, including Seonmin Will Heo, Eunseo Kang, and James Banovetz.

**Definition 3**

Let $S$ be the sample space, and let $A$ and $B$ be events defined on $S$. Then:

- $A$ is a subset of $B$, written $A \subset B$, if every element of $A$ is also an element of $B$.
- The event with no outcomes, $\emptyset = \{\}$, is called **empty set**.
- The **union** of $A$ and $B$, denoted $A \cup B$, is the collection of all outcomes that are in either $A$ or $B$ (or both).
- The **intersection** of $A$ and $B$, denoted $A \cap B$, is the collection of outcomes that are in both $A$ and $B$.
- The **complement** of $A$, denoted $A^c$, is the set of all outcomes in $S$ that are not in $A$.
- The events $A$ and $B$ are **disjoint** if they have no outcomes in common: $A \cap B = \emptyset$.
- The events $A_1, A_2, ...$ are **pairwise disjoint** if $A_i \cap A_j = \emptyset$ for all $i \neq j$.
- The events $A_1, A_2, ..., A_n$ are a **partition** of $S$ if they are pairwise disjoint and their union is $S$ ($\cup_{i=1}^{\infty} A_i = S$).

The following theorem summarizes some properties of set operations.

**Theorem 1**

For any events $A$, $B$, $C$, and $\{E_i\}_{i=1}^{\infty}$ defined on the sample space $S$:

- **Commutativity** $A \cup B = B \cup A$
  $A \cap B = B \cap A$
- **Associativity:** $A \cup (B \cup C) = (A \cup B) \cup C$
  $A \cap (B \cap C) = (A \cap B) \cap C$
- **Distributive Laws:**

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \text{ and } A \cap \left(\bigcup_{i=1}^{\infty} E_i\right) = \bigcup_{i=1}^{\infty} \left(A \cap E_i\right)$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \text{ and } A \cup \left(\bigcap_{i=1}^{\infty} E_i\right) = \bigcap_{i=1}^{\infty} \left(A \cup E_i\right)$$

- **De Morgan's Laws:** $(A \cup B)^c = A^c \cap B^c$ and $\left(\bigcup_{i=1}^{\infty} E_i\right)^c = \bigcap_{i=1}^{\infty} E_i^c$
  $(A \cap B)^c = A^c \cup B^c$ and $\left(\bigcap_{i=1}^{\infty} E_i\right)^c = \bigcup_{i=1}^{\infty} E_i^c$

Now we turn to a concept that is relevant for defining probability: the sigma algebra.

**Definition 4** (Sigma algebra)

Given a sample space $S$, a collection of subsets of $S$ is called a **$\sigma$-algebra (sigma algebra)**, denoted by $\mathcal{B}$, if it satisfies the following three properties:

1. $\emptyset \in \mathcal{B}$ (the empty set is an element of $\mathcal{B}$)
2. If $A \in \mathcal{B}$, then $A^C \in \mathcal{B}$ ($\mathcal{B}$ is closed under complementation), and
3. If $A_1, A_2, ... \in \mathcal{B}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{B}$ ($\mathcal{B}$ is closed under countable unions).

These properties also imply the following useful facts:

- $S \in \mathcal{B}$ (since $\emptyset \in \mathcal{B}$ and $S = \emptyset^C$)
- $\mathcal{B}$ is closed under countable intersections: $\cap_{i=1}^{\infty} A_i \in \mathcal{B}$ (by De Morgan's Law and properties 2 and 3).

Why do we need $\sigma$-algebras to define probability? The reason is that in cases involving uncountably infinite sample spaces, it becomes necessary to restrict the set of allowable events. We want to ensure that we work only with the *measurable* sets, those for wich areas are well-defined. While this is a technicality that rarely affects econometrics in practice[2], it is important to be familiar with the terminology, as it is frequently used in probability theory.

**Example 1**

Consider the sample space $S = \{1, 2, 3\}$.

One $\sigma$-algebra is the trivial $\sigma$-algebra, given by $\mathcal{B} = \{\emptyset, S\}$.

The sigma algebra we will typically work with is the power set of $S$: $\mathcal{B} = \{\text{all subsets of } S\}$.

Since $S$ has $n = 3$ elements, the power set contains $2^3 = 8$ subsets, the collection of which forms the sigma algebra:

$$\mathcal{B} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$$

[2]These technicalities do not arise when $S$ is finite or countable.

### Notions of Probability

Probability is a concept that lies at the heart of both economic theory and econometrics. It defines the mathematical language we use to model uncertainty, variability, and randomness. Before introducing formal definitions of probability, we begin with the concept of a measure.

**Definition 5** (Measure)

Let $S$ be a sample space with associated $\sigma$-algebra $\mathcal{B}$. A **measure** $\mu$ on $S$ with $\sigma$-algebra $\mathcal{B}$ is a function $\mu : \mathcal{B} \to [0, \infty)$ such that:

1. The measure of the empty set is zero: $\mu(\emptyset) = 0$, and
2. $\mu$ is countably additive: $\mu\big(\bigcup_{i=1}^{\infty} E_i\big) = \sum_{i=1}^{\infty} \mu(E_i)$ for any $E_1, E_2, \dots \in \mathcal{B}$ where $E_i \cap E_j = \emptyset \ \ \forall i \neq j$.

**Definition 6** (Probability Function)

Given a sample space $S$ and an associated sigma algebra $\mathcal{B}$, a **probability function** is a measure $\mathbb{P}$ with domain $\mathcal{B}$ that satisfies the following **Axioms of Probability**:

1. $\mathbb{P}(A) \geq 0$ for all $A \in \mathcal{B}$,
2. $\mathbb{P}(S) = 1$, and
3. If $A_1, A_2, \ldots$ are pairwise disjoint, then $\mathbb{P}\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

This is known as the axiomatic definition of probability. Under this definition, any function $\mathbb{P}$ that satisfies these axioms is considered a probability function. Note that probability is a function from the space of events to the non-negative real numbers.

From these axioms, we can derive several properties of the probability function.

**Theorem 2: Properties of Probability**

Let $\mathbb{P}$ be a probability function and let $A, B$ be sets in $\mathcal{B}$, then:

- $\mathbb{P}(\emptyset) = 0$
- $\mathbb{P}(A) \leq 1$
- $\mathbb{P}(A^c) = 1 - P(A)$
- $\mathbb{P}(B \cap A^c) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ (inclusion-exclusion principle)
- If $A \subseteq B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$

We now present two important inequalities derived from these properties, which are frequently used to bound probabilities.

**Theorem 3**

Let $\mathbb{P}$ be a probability function and let $A, B, E_1, ...$ be sets in $\mathcal{B}$, then:

- **Bonferroni's inequality:** $\mathbb{P}(A \cap B) \geq P(A) + P(B) - 1$
- **Boole's inequality:** $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$
  $\mathbb{P}\left(\cup_{i=1}^{\infty} E_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(E_i)$

**Example 2**

Consider an experiment consisting of tossing a fair coin. Then the sample space is $S = \{H, T\}$, where $H$ denotes heads and $T$ denotes tails.

By "fair", we mean that we would expect the event of a heads is to be as likely as the event of a tails. Thus, a reasonable probability function would satisfy:

$$\mathbb{P}(\{H\}) = \mathbb{P}(\{T\})$$

Using the axioms of probability:

1. $\mathbb{P}(S) = \mathbb{P}(\{H\} \cup \{T\}) = 1$
2. $\mathbb{P}(\{H\}) + \mathbb{P}(\{T\}) = 1$ (since $T$ and $H$ are disjoint and thus $\mathbb{P}(\{H\} \cup \{T\}) = \mathbb{P}(\{H\}) + \mathbb{P}(\{T\})$).
3. $\mathbb{P}\big(\{H\}\big) = \mathbb{P}\big(\{T\}\big) = \dfrac{1}{2} \geq 0$

In cases like the one above, where all outcomes in $S$ are equally likely, probabilities of events can be calculated by simply counting the number of outcomes in the event.

Suppose $S = \{s_1, ..., s_N\}$ is a finite sample space. Saying that all outcomes are equally likely means that $\mathbb{P}(\{s_i\}) = \frac{1}{N}$ for every outcome $s_i$. Then, for any event $A$, the probability of $A$ is:

$$\mathbb{P}(A) = \sum_{s_i \in A} \mathbb{P}(\{s_i\}) = \sum_{s_i \in A} \frac{1}{N} = \frac{|A|}{|S|}$$

where $|A|$ denotes the number of elements in the set $A$.

## Counting

In many probability calculations, it is useful to count the number of individual outcomes. Counting methods are specially useful when constructing probability assignments on finite sample spaces.

We begin with the counting rule, which shows how to compute the number of outcomes when an experiment consists of multiple stages.

**Theorem 4: Counting Rule**

If an experiment consists of $k$ separate stages, where the $i^{th}$ stage has $n_i$ possible outcomes for $i = 1, ..., k$, then the total number of possible outcomes is: $n_1 \times n_2 \times \cdots \times n_k$.

This rule is also known as the Fundamental Theorem of Counting.

**Example 3**

Suppose license plates are formed using three letters (A-Z) followed by four numerical digits (0-9). If repeated letters and digits are allowed, how many distinct license plates are possible?

$$26 \times 26 \times 26 \times 10 \times 10 \times 10 \times 10 \approx 175 \text{ million}$$

We now introduce a useful notation:

**Definition 7**

The **factorial** of a natural number $n \in \mathbb{N}$ is the product of all positive integers less than or equal to $n$:

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1 = \prod_{i=1}^{n} i$$

Let us now consider the problem of sampling from a finite set. The number possible outcomes depends on two factors:

1. Can elements be repeated?
2. Does the order matter?

There are four canonical cases, summarized in the table below:

Table 1: Number of possible arrangements of size $r$ from $n$ objects

| | Without Replacement | With Replacement |
|---|---|---|
| Ordered | $P_r^n = \frac{n!}{(n-r)!}$ | $n^r$ |
| Unordered | $C_r^n = \binom{n}{r} = \frac{n!}{(n-r)!r!}$ | $\binom{n+r-1}{r} = \frac{(n+r-1)!}{(n-1)!r!}$ |

Let us explore each case with examples:

1. **Ordered, Without Replacement** (Permutations)

$$P_r^n = \frac{n!}{(n-r)!}$$

*Example:* Padlock "combinations". A padlock has 40 digits and requires three distinct digits in the correct order to unlock. How many possible padlock "combinations" are there?

$$40 \times 39 \times 38 = \frac{40!}{37!} = 59,280$$

2. **Ordered, With Replacement** This corresponds to the fundamental theorem of counting, where each stage has the same number of options.

$$n^r$$

*Example:* Some states issue truck license plates with only six numerical digits (0-9), allowing for repetition. How many variations of these license plates are possible?

$$10^6 = 1,000,000$$

3. **Unordered, Without Replacement** (Combinations)

$$C_r^n = \binom{n}{r} = \frac{n!}{(n-r)!r!}$$

*Example:* Suppose you have 5 positions in your PhD program, but 30 equally qualified applicants. How many different incoming classes could you select?

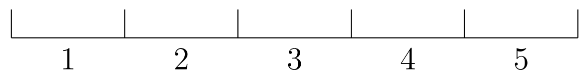
$$\binom{30}{5} = \frac{30!}{(25!)(5!)} = 142,506$$

4. **Unordered, With Replacement**

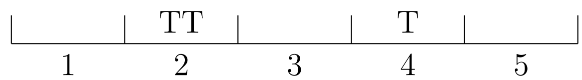
$$\binom{n+r-1}{r} = \frac{(n+r-1)!}{(n-1)!r!}$$

*Example:* Assume we have five potential job sites, and three identical trucks, where multiple trucks can go to the same site. Let us visualize this as placing trucks into bins:

- Bins: Consider the 5 sites as "bins," numbered 1–5 ($n = 5$).

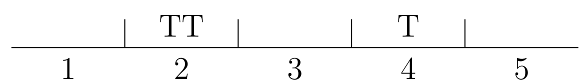



- Trucks: Identical units assigned to bins ($r = 3$).




  This corresponds to two trucks at site 2 and one at site 4 (alternatively, this might be seen as the result of drawing two 2s and one 4).

- Consider each bin "wall" and each truck as an element to be ordered. Note that the first and last walls are "immobile", so we will not consider them:




  We may represent this as a sequence of trucks ($T$) and dividers or walls ($W$) between bins. Thus, this corresponds to the ordering $WTTWWTW$.

- We have seven total positions: 3 trucks + 4 dividers. If they were distinct elements, we would have 7! possibilities. Trucks and dividers are indistinguishable among themselves. Thus the number of distinct assignments is:

$$\frac{7!}{4!3!} = \binom{7}{3}$$

  which corresponds to our formula for unordered, with replacement, when we have five objects, picking three.

Since we are already discussing methods of counting and sampling, it is worth briefly introducing two methods often used in econometrics:

1. **Monte Carlo simulations**: Monte Carlo methods refer to algorithms that involve repeated random sampling to estimate numerical results. *Example:* Suppose we want to approximate the distribution of the sum of two fair dice. Each die has six faces, each

with equal chance of occurrence (probability of 1/6). Thus we simulate the following process:

- Randomly generate two integers between 1 and 6
- Repeat this process 10,000 times
- Plot the histogram of the resulting sums

As the number of simulations increases, the empirical distribution will be closer to the theoretical one.

2. **Bootstrapping**: Bootstrapping is a random sampling method used to estimate a metric or run a test by sampling *with replacement* from the observed data. It is very often used to compute standard errors of a regression coefficient. *Example:* Suppose we have a dataset with 5,000 observations. This is the process to estimate a standard error:
   - Sample the same number of observations ($N = 5000$) from our sample with replacement
   - Run the regression on this bootstrapped sample and record the standard errors
   - Repeat this process several times, for example 10,000 times
   - Compute the mean of the 10,000 standard errors to obtain the bootstrapped standard error.

### Conditional Probabilities and Independence

In many applications, we are interested in the relationship between two events. For example, suppose we want to understand the relationship between wages and education. We collect data on a randomly selected group of people, classifying them based on college education status (college, $C$, or no college education, $N$), and wage level (high, $H$, or low, $L$, wage). This information is presented in the table below:

| | C | N | Total |
|---|---|---|---|
| H | 10 | 6 | 16 |
| L | 8 | 26 | 34 |
| Total | 18 | 32 | 50 |

Using this information, we may want to answer questions like: 1) what is the probability that a person has college education and a low wage? or 2) if a person has college education, what is the probability that they have low wage? These questions are not equivalent.

To answer the second question, we introduce the concept of conditional probability.

**Definition 8** (Conditional Probability)

If $A$ and $B$ are events in $S$, and $\mathbb{P}(B) > 0$, then the **conditional probability** of $A$ given $B$, denoted $\mathbb{P}(A|B)$, is defined as

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Given $B \in \mathcal{B}$ such that $\mathbb{P}(B) \neq 0$, $\mathbb{P}(\cdot|B) : \mathcal{B} \to [0, \infty)$ is a probability measure on $S$ with $\sigma$-algebra $\mathcal{B}$.

**Example 4**

Suppose we toss a fair six-sided die. What is the probability that we observe a 1, given that we observe an odd number?

$$\begin{aligned}
\mathbb{P}(\text{odd}) &= 1/2 && \text{(three odds out of six)} \\
\mathbb{P}(\text{1 and an odd}) &= 1/6 && \text{(one 1 out of six total)} \\
\mathbb{P}(\text{one}|\text{odd}) &= \frac{\mathbb{P}(\text{one and an odd})}{\mathbb{P}(\text{odd})} && \text{(by def. of cond. prob.)} \\
\mathbb{P}(\text{one}|\text{odd}) &= \frac{1/6}{1/2} = 1/3
\end{aligned}$$

**Definition 9** (Statistical Independendence)

Two events $A$ and $B$ in $S$ are said to be **independent** if and only if we have one of three equivalent conditions:

- $\mathbb{P}(A|B) = \mathbb{P}(A)$
- $\mathbb{P}(B|A) = \mathbb{P}(B)$
- $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$

An important relationship can be derived from the partitioning theorem.

**Theorem 5: Law of Total Probability**

Let $A_1, A_2, ...$ be a partition of the sample space $S$, and that $\mathbb{P}(A_i) > 0$ for each $i$. If $B$ is an event, then

$$\mathbb{P}(B) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)\mathbb{P}(B|A_i)$$

We now present a famous result credited to Reverend Thomas Bayes, which applies the definition of conditional probability.

**Theorem 6: Bayes' Rule**

Let $A_1, A_2, ...$ be a partition of the sample space $S$, and let $A$ and $B$ be events in a sample space $S$. Then:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{j=1}^{\infty} \mathbb{P}(B|A_j)\mathbb{P}(A_j)}$$

## References

Casella, G. and Berger, R. (2002). *Statistical inference.* Chapman and Hall/CRC, 2nd edition.

Hansen, B. (2022). *Probability and statistics for economists.* Princeton University Press.