Statistics Lecture

Claudio Ortelli [1]

[1]Finance Institute
USI

Advanced Learning and Research Institute, 2012

Definition

A random experiment is an experiment whose outcome can not predicted with certainty.

Examples

- Observing if a system component is functioning properly or has failed at a given point in time in the future.
- Determining the execution time of a program.
- Determining the response time of a server request.

Definition

The result of the experiment is called the outcome of the experiment. The totality of the possible outcomes of a random experiment is called the sample space of the experiment and it will be donoted by the letter S.

- The definition of the sample space S is determined by the experiment and the purpose for which the experiment is carried out. When observing the status of two components of a running system, it may be sufficient to know if zero, one or two components have failed without having to exactly identify which component has failed.
- We classify the sample spaces w.r.t. the number of elements they contain.
    - *Finite sample* space: the set of possible outcomes of the experiment is finite;
    - *Countably infinite* sample space: the outcomes of the experiment are in a one-to-one relationship with $\mathbb{N}$;
    - Otherwise the sample space is called *uncountable* or *nondenumerable*.
- A finite or countably finite sample space is called a *discrete* sample space. *Continuous* sample spaces, such as all the points on a line or interval are examples of uncountables sample spaces.

Definition

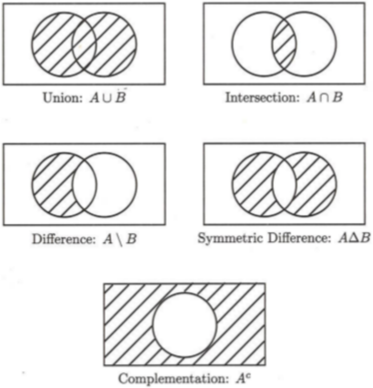
Given a random experiment and the corresponding sample space S, a collection of certain outcomes is called an event E. E is a subset of the sample space $E \subset S$. Equivalently, any statement of conditions identifying E is called an event.

Example

In the random experiment "Toss of a die", we define the sample space $S = \{1, \ldots, 6\}$. The event $E = \{$The outcome is an even number$\}$ is equivalent to $E = \{2, 4, 6\}$.

Definition

Given a random experiment with sample space S we call a single performance of the experiment a *trial*.



Union: $A \cup B$
Intersection: $A \cap B$
Difference: $A \setminus B$
Symmetric Difference: $A \Delta B$
Complementation: A^c

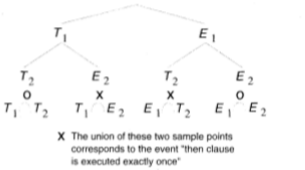


Figure 1.8. Tree diagram of a sequential sample space

The set of all leaves of the tree is the sample space of interest.

Basic definitions

Two-dimensional sample space

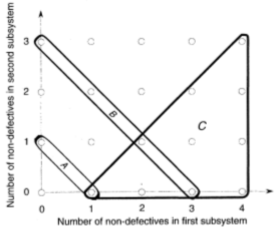


Figure 1.9. A two-dimensional sample space

1. A :={"the system has exactly one non-defective component"}
2. B :={"the system has exactly three non-defective components"}
3. C :={"the first subsystem has more non-defective components than the second subsystem"}

- Let E be an event of S and let denote the outcome of a specific trial by s. If $s \in E$ we then say that the event E has *occured*. Only one outcome $s \in S$ can occur on any trial. However, every event including s will occur.
- Let E, E_1 and E_2 be events. We define the event $\bar{E}$ to be *the complement* of E, i.e. $\bar{E} = \{s \in S \mid s \notin E\}$.
- Let E_1 and E_2 be two events. We define the event
    - $E_1 \cup E_2 := \{s \in S \mid s \in E_1 \text{ or } s \in E_2\}$ *the union* of E_1 and E_2.
    - $E_1 \cap E_2 := \{s \in S \mid s \in E_1 \text{ and } s \in E_2\}$ *the intersection* of E_1 with E_2.
    - $E_1 \setminus E_2 := \{s \in S \mid s \in E_1 \text{ and } s \notin E_2\}$ the difference of $E_1$and E_2.
- If $E_1 \cap E_2 = \emptyset$ we say that the two events E_1 and E_2 are *mutually exclusive* or *disjoint*.
- We denote by $\mid E \mid$ *the cardinality* of E, i.e. the number of elements (outcomes) in E.

Definition

Let S denotes the sample space of a given experiment and $\mathscr{F}$ a collection of events. We say that $\mathscr{F}$ is an *algebra* over S if the following two conditions are fulfilled:

1. S must be an element of $\mathscr{F}$, i.e. $S \in \mathscr{F}$.
2. If $E_1 \in \mathscr{F}$, $E_2 \in \mathscr{F}$, the sets $E_1 \cup E_2$, $E_1 \cap E_2$ must also belong to $\mathscr{F}$.
3. If $E_1 \in \mathscr{F}$, $E_2 \in \mathscr{F}$, $E_1 \setminus E_2$ must also belong to $\mathscr{F}$.

Interpretation: an algebra $\mathscr{F}$ over S is a family of subsets of S which is closed with respect to the three binary operators $\cup, \cap$ and $\setminus$.
Remark: condition 3 of the previous definition can be replaced by the following equivalent condition:

- 3^{bis} If $E \in \mathscr{F}$ then $\overline{E} \in \mathscr{F}$.

Example

a) The collection given by $\mathscr{F} = \{S, \emptyset\}$ is an algebra (the so called *trivial* algebra over S).
b) The collection $\mathscr{F} = \{E, \bar{E}, S, \emptyset\}$ is the algebra generated by $E \subset S$.
c) The power set of S, denoted by $\mathscr{P}(S)$, is defined to be the collection consisting of all subsets of S (including the empty set $\emptyset$).

Exercise

1. *Show by mathematical induction that if S is a finite set with $\mid S \mid = n$ elements, then the power set of S contains $\mid \mathscr{P}(S) \mid = 2^n$ elements.*
2. *Show that condition 3^{bis} is equivalent to condition 3.*
3. *Given $S = \{1, 2, 3, 4, 5, 6\}$ construct at least two different algebras on S.*
4. *De Morgan's law. Let A and B two events. Show that*

$$\overline{A \cup B} = \overline{A} \cap \overline{B} \text{ and } \overline{A \cap B} = \overline{A} \cup \overline{B}.$$

Exercise

1. *De Morgan's law. Let I be a non empty, possibly uncountable set and $A_i, i \in I$ a family of sets indexed by I. Show that*

$$\overline{\bigcup_{i\in I} A_i} = \bigcap_{i\in I} \overline{A}_i \text{ and } \overline{\bigcap_{i\in I} A_i} = \bigcup_{i\in I} \overline{A}_i.$$

Definition

The indicator of the set $A \subseteq S$ is the function on S given by

$$1_A(s) = \begin{cases} 1 & \text{if } s \in A \\ 0 & \text{if } s \notin A. \end{cases}$$

Interpretation: the function1_A "indicates" wheter A occurs.

Exercise

Prove the following equalities:

$1_{A\cup B} = max\{1_A, 1_B\}$; $1_{A\cap B} = min\{1_A, 1_B\} = 1_A 1_B$

$1_{\overline{A}} = 1 - 1_A$; $1_{A\triangle B} = \mid 1_A - 1_B \mid$

Definition

Let S denotes the sample space of a given experiment and $\mathscr{G}$ a collection of events. We say that $\mathscr{G}$ is a $\sigma-$*algebra* over S if

1. $\mathscr{G}$ is an *algebra* over S.
2. If $E_n \in \mathscr{G}$, $n = 1, 2, \ldots$, then

$$\bigcup_{n=1}^{\infty} E_n \in \mathscr{G}.$$

Interpretation: a $\sigma-$algebra $\mathscr{G}$ over S is a family of subsets of S which is closed with respect to

- the difference operator $\setminus$ (or, equivalently, complementation),
- the *countable* union and intersection of its elements.

Theorem

Let $\mathscr{E}$ be a collection of subsets of S. Then there are a smallest algebra $\alpha(\mathscr{E})$ and a smallest $\sigma-$algebra $\sigma(\mathscr{E})$ containing all the sets that are in $\mathscr{E}$.

Proof.

$\mathscr{P}(S)$ is a $\sigma-$algebra on S. Therefore it exists at least one $\sigma-$algebra and one algebra containing $\mathscr{E}$. We define $\alpha(\mathscr{E})$ (or $\sigma(\mathscr{E})$) to consist of all sets that belong to every algebra (or $\sigma-$algebra) containing $\mathscr{E}$. It is easy to verify that this system is an algebra (or $\sigma - algebra$) and indeed the smallest. $\square$

Examples

1. Let $\mathscr{C}$ be the family of open intervals on the real line, i.e.

$$\mathscr{C} := \{(a,b) \mid a,b \in \mathbb{R} \text{ and } a < b\}.$$

$\mathscr{B}(\mathbb{R}) := \sigma(\mathscr{C})$ is called the Borel sigma algebra over $\mathbb{R}$.

2. Let $f : S \to \mathbb{R}$ be a real valued function. The family of preimages:

$$\{f^{-1}(B) : B \in \mathscr{B}(\mathbb{R})\}$$

is a $\sigma-$algebra on S, the $\sigma-$algebra generated by f [denoted $\sigma(f)$].

Definition

The pair $(S,\mathscr{G})$ where S is the sample space and $\mathscr{G}$ a $\sigma-$algebra on S is called a measurable space and the elements of $\mathscr{G}$ are called events.

Remark: a subset A of the sample space S is an event if and only if $A \in \mathscr{G}$.

Example

Suppose we toss a die once. We choose $S = \{1,2,3,4,5,6\}$ and $\mathscr{G} := \{S,\emptyset,\{1,2,3\},\{4,5,6\}\}$. The subset $A = \{2,4,6\}$ of S is *not* an event of $\mathscr{G}$. Let define the sigma algebra $\mathscr{H} := \sigma(\{\{1,2,3\},A\})$ and the new measurable space $(S,\mathscr{H})$. $A \in \mathscr{H}$ so that A is now an event.

Exercise

Complete the sigma algebra $\mathscr{H}$ by enumerating its elements.

Definition

Let $(S, \mathscr{H})$ be a measurable space. A probability measure or probability law is a positive real-valued function $P : \mathscr{H} \rightarrow [0, 1]$ such that the following axioms hold

1. $P(S) = 1$.
2. For every sequence $\{E_n\}_{n \in \mathbb{N}}$ of pairwise disjoint events ($E_i \cap E_j = \emptyset$, $\forall i \neq j$) it must hold

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

Remark: a probability law is a function defined on a sigma algebra. The *arguments* of P are *events*, i.e. *subsets* of S and *elements* of $\mathscr{H}$.
A probability law has many useful relations, see Trivedi pp. 15-16.

Example

Probability of union of events.

If $A_1, A_2, \dots A_n$ are any events, then

$$
\begin{aligned}
P(\bigcup_{i=1}^{n} A_i) &= P(A_1 \cup A_2 \cup \cdots \cup A_n) \\
&= \sum_i P(A_i) - \sum_{1 \le i < j \le n} P(A_i \cap A_j) \\
&\quad + \sum_{1 \le i < j < k \le n} P(A_i \cap A_j \cap A_k) + \cdots \\
&\quad + (-1)^{n-1} P(A_1 \cap A_2 \cap \cdots \cap A_n),
\end{aligned}
$$

where the successive sums are over all possible events, pairs of events, triples of events, and so on.

$$
\begin{aligned}
P(\bigcup_{i=1}^{n} A_i) &= P(A_1) + P(\overline{A}_1 \cap A_2) + P(\overline{A}_1 \cap \overline{A}_2 \cap A_3) + \cdots \\
&\quad + P(\overline{A}_1 \cap \overline{A}_2 \cap \cdots \cap \overline{A}_{n-1} \cap A_n). \qquad (1.4)
\end{aligned}
$$

Exercise

Using the properties of P prove equality 1.4