# Probability theory

## Probability and distributions

Francesco Corona

Department of Computer Science
Federal University of Ceará, Fortaleza

# Introduction

## Probability theory

# Introduction

Many kinds of studies can be characterised as (repeated) experiments

- Under (essentially) identical conditions
- More or less a standard procedure

- Medical sciences: Effect of a drug that is to be provided
- Economic sciences: Prices of some commodities in time
- Agronomic sciences: Effect that a fertiliser has on yield

To get information about such phenomena we perform an experiment

- Each experiment ends with an **outcome**

The outcome can only be predicted with uncertainty

# Introduction (cont.)

Imagine that we have such an experiment

Experiments are s.t. we can describe a collection of every possible outcomes

- Already before the performance of the experiment

- If the experiment can be repeated under 'identical' conditions, then we denote it a **random experiment**
- We denote the collection of every possible outcome of a random experiment the **sample space**, or the **experimental space**

# Introduction (cont.)

## Example

**Toss of a coin**

- Let the outcome 'tails' be denoted by $\mathtt{T}$
- Let the outcome 'heads' be denoted by $\mathtt{H}$

The toss of this coin is an example of a random experiment

- The outcome of the experiment is one of two symbols, $\mathtt{T}$ and $\mathtt{H}$
- The sample space is the collection of these two symbols, $\{\mathtt{T}, \mathtt{H}\}$

We accept that the coin can be repeatedly tossed under the same conditions

■

# Introduction (cont.)

## Example

**Cast of two coloured dice**

- Let the outcome be an ordered pair
- `(Number of spots on red die, Number of spots on white die)`

If we accept that these two dice can be repeatedly cast under the same exact conditions, then the cast of these dice is an example of a random experiment

- The sample space is the set of 36 ordered pairs
- $\{(1,1), \cdots, (1,6), (2,1), \cdots, (2,6), \cdots, (6,6)\}$

■

# Introduction (cont.)

Let $\mathcal{C}$ denote the sample space, let $c$ denote an element of $\mathcal{C}$

- Let $C$ represent a collection of elements of $\mathcal{C}$

After performing the experiment, let the outcome be in $C$

$\rightsquigarrow$ We say that the **event** $C$ has occurred

# Introduction (cont.)

Imagine of having $N$ repeated executions of the random experiment

- We can count the number $f$ of times that event $C$ occurred (throughout the $N$ executions)
- The **number of occurrences** (the **frequency**)

Ratio $f/N$ is the **relative frequency** of event $C$ (in the $N$ experiments)

# Introduction (cont.)

Relative frequency can be quite erratic a quantity, for small values of $N$

- As $N$ increases, we could start associating with event $C$ a number, $p$
- $p$ approximates the number about which relative frequency stabilises

We can start coming up with an interpretation of number $p$

- It is that number that relative frequency of event $C$ will approximate

# Introduction (cont.)

### Remark

Summarising, we cannot predict the outcome of a random experiment

We can approximate the relative frequency with which it will be in $C$

- $\rightsquigarrow$ For large $N$

# Introduction (cont.)

The number $p$ associated with event $C$ has received various names

- The **probability** of event $C$
- The **probability measure** of $C$
- The **probability** that the outcome (of random experiment) is in $C$

The context suggests appropriate choices of terminology

# Introduction (cont.)

## Example

**Cast of two coloured dice**

- Let $\mathcal{C} = \{(1,1), \cdots, (1,6), (2,1), \cdots, (2,6), \cdots, (6,6)\}$
- Let $C$ be the collection of all pairs with sum equal 7

$$\{(1,6), (2,5), (3,4), (4,3), (5,2), (1,6)\}$$

Say we cast $N = 1K$ times, let $f = 150$ be the frequency of sum equal 7

$$f/N = \frac{150}{1000} = 0.15$$

We can associate with event $C$ a number $p$ that is close to $f/N$

$\rightsquigarrow$ $p$ would define the probability of event $C$

# Introduction (cont.)

## Remark

The provided interpretation of probability is referred to as frequentist

- It is subjected to the fact that an experiment can be repeated
- (Under the same 'identical' conditions)

The interpretation of probability can be extended to other situations

- By treating it as a subjective (rational?) measure of belief

The math properties of probability are consistent with either of the meanings

- Our development won't depend up on the interpretation

# Introduction (cont.)

The course is about defining mathematical models for random experiments

First a model has to be developed with the theory set in place, then we can start making inferences (draw conclusions) about the random experiment

Building such a model requires a theory of probability

- A satisfactory theory of probability is to be developed
- We shall use the concepts of sets and functions of sets

# Set theory

## Probability theory

# Set theory

Typically, the concept of a **set**/**collection** of objects can be left undefined

- Yet, we need to minimise the risk of misunderstanding
- What collection of objects is under study?
- Particular sets can (must) be described

## Example

$$\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

The set of the first 10 non-negative integers is sufficiently well described to clarify that numbers $2/7$ and $-5$ are not in it, while number 7 is in it

# Set theory (cont.)

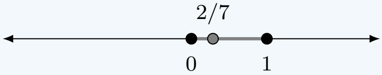
If an object belongs to a set, it is said to be an **element** of that set

## Example

Let $C$ denotes the set of real numbers $x$ such that $0 \leq x \leq 1$

- $2/7$ is an element of set $C$

The fact that $2/7$ is an element of set $C$ is denoted by $2/7 \in C$




More generally, $c \in C$ means that $c$ is an element of set $C$

# Set theory (cont.)

Our concern is on sets that are, typically, sets of numbers

- We use the the language of '**sets of points**'
- More convenient than 'sets of numbers'

In analytical geometry, to each point on the line (onto which origin- and unit-point have been selected) corresponds one and only one number, $x$

- ⥀ To each number $x$ corresponds one and only one point on the line

The 1-to-1 correspondence between numbers and points on a line allows us to speak of the 'point $x$' meaning the 'number $x$' (without ambiguity)

# Set theory (cont.)

On a planar rectangular system of coordinates and with $x$ and $y$ numbers, to each symbol $(x, y)$ there corresponds one and only one point in the plane

- To each point in the plane corresponds one and only one symbol $(x, y)$

Because of the one-to-one correspondence between numbers and points, we can speak of the 'point $(x, y)$' instead of the 'ordered number-pair $x$ and $y$'

# Set theory (cont.)

The same vocabulary can be used with a rectangular system of coordinates in a space of three, four, five, six, seven, eight, nine, ten, ... or more dimensions

The 'point $(x_1, x_2, \dots, x_n)$' means the ordered number $n$-tuple $x_1, x_2, \dots, x_n$

# Set theory (cont.)

The bottom line: When describing sets, we speak of sets of points

- Sets whose elements are points

And, we want to be precise in describing our sets

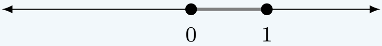
- ... to avoid any ambiguity

# Set theory (cont.)

## Example

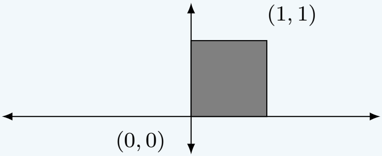
Notation $C = \{x : 0 \leq x \leq 1\}$

⇝ '$C$ is the mono-dimensional set of points $x$ such that $0 \leq x \leq 1$'




Notation $C = \{(x, y) : 0 \leq x, y \leq 1\}$

⇝ '$C$ is the bi-dimensional set of points $(x, y)$ that are internal to, or on the frontier of, a square whose facing vertices are $(0, 0)$ and $(1, 1)$'



# Set theory (cont.)

## Remark

When the dimensionality of sets is clear, no need to make reference to it

# Set theory (cont.)

A set $C$ is said to be **countable** when either $C$ is finite, or it has as many elements as there are positive integers (aka, counting or natural numbers $\mathcal{Z}$)

## Example

- Sets $C_1 = \{1, 2, \dots, 100\}$ and $C_2 = \{1, 3, 5, 7, \dots\}$ are countable sets
- The interval of real numbers $(0, 1]$ is not a countable set

# Set theory (cont.)

Some definitions, together with some illustrative examples

⇝ An elementary algebra of sets

# Set theory (cont.)

## Definition

*If each element of some set $C_1$ is also an element of some set $C_2$, then we say that set $C_1$ is a **subset** of set $C_2$*

- *This is denoted by writing $C_1 \subset C_2$*

*If $C_1 \subset C_2$ and also $C_2 \subset C_1$, then sets $C_1$ and $C_2$ have the same elements*

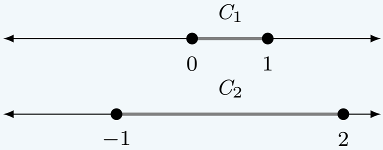
- *This is denoted by writing $C_1 = C_2$*

# Set theory (cont.)

## Example

Define the sets

- $C_1 = \{x : 0 \leq x \leq 1\}$
- $C_2 = \{x : -1 \leq x \leq 2\}$




Each element of set $C_1$ is also an element of set $C_2$

Mono-dimensional set $C_1$ is a subset of mono-dimensional set $C_2$

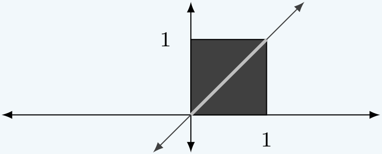
⇝ $C_1 \subset C_2$

# Set theory (cont.)

## Example

Define the sets

- $C_1 = \{(x, y) : 0 \leq x = y \leq 1\}$
- $C_2 = \{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1\}$




The elements of $C_1$ are points on one diagonal of square $C_2$

⇝ $C_1 \subset C_2$

# Set theory (cont.)

## Definition

*If a set $C$ has no elements, it is said that $C$ is the* ***null set***

- *This is denoted by writing $C = \emptyset$*

# Set theory (cont.)

## Definition

*The set of all elements that are in at least one of the sets $C_1$ and $C_2$*

⇝ *The* ***union*** *of $C_1$ and $C_2$*

- *The union of $C_1$ and $C_2$ is denoted by writing $C_1 \cup C_2$*

*The set of all elements are in at least one of the sets $C_1, C_2, C_3, \cdots$*

⇝ *The union of the sets $C_1, C_2, C_3, \cdots$*

- *If an infinite number of sets is considered, their union is indicated by*

$$C_1 \cup C_2 \cup C_3 \cup \cdots = \bigcup_{j=1}^{\infty} C_j$$

- *If a finite number of sets is considered, their union is indicated by*

$$C_1 \cup C_2 \cup C_3 \cup \cdots \cup C_k = \bigcup_{j=1}^{k} C_j$$

*We refer to the union $\bigcup_{j=1}^{\infty} C_j$ as a* ***countable union***

# Set theory (cont.)

## Example

Define the sets

- $C_1 = \{x : x = 8, 9, 10, 11 \text{ or } 11 < x \leq 12\}$
- $C_2 = \{x : x = 0, 1, \cdots, 10\}$

Then,

$$\begin{aligned} C_1 \cup C_2 &= \{x : x = 0, 1, \cdots, 8, 9, 10, 11 \text{ or } 11 < x \leq 12\} \\ &= \{x : x = 0, 1, \cdots, 8, 9, 10 \text{ or } 11 \leq x \leq 12\} \end{aligned}$$

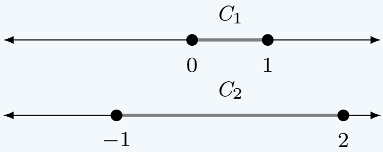
■

# Set theory (cont.)

## Example

Define the sets

- $C_1 = \{x : 0 \leq x \leq 1\}$
- $C_2 = \{x : -1 \leq x \leq 2\}$




Then,

$$C_1 \cup C_2 = C_2$$

# Set theory (cont.)

## Example

Let $C_2 = \emptyset$, then $C_1 \cup C_2 = C_1$ (for every set $C_1$)

■

## Example

For every set $C$, $C \cup C = C$

■

# Set theory (cont.)

## Example

Let

$$C_k = \left\{ x : \frac{1}{k+1} \leq x \leq 1 \right\}, \quad k = 1, 2, 3, \cdots$$

Then,

$$\bigcup_{k=1}^{\infty} C_k = \{x : 0 < x \leq 1\}$$

The number zero is not in any of the sets $C_k$

- ⇝ Number zero is also not in their union set

# Set theory (cont.)

## Definition

*The set of all elements that are in each of the sets* $C_1$ *and* $C_2$

⇝ *The* ***intersection*** *of* $C_1$ *and* $C_2$

- *The intersection of* $C_1$ *and* $C_2$ *is denoted by writing* $C_1 \cap C_2$

*The set of all elements that are in each of the sets* $C_1, C_2, C_3, \cdots$

⇝ *The intersection of sets* $C_1, C_2, C_3, \cdots$

- *If an infinite number of sets is considered, intersection is indicated by*

$$C_1 \cap C_2 \cap C_3 \cap \cdots = \bigcap_{j=1}^{\infty} C_j$$

- *If a finite number of sets is considered, intersection is indicated by*

$$C_1 \cap C_2 \cap C_3 \cap \cdots \cap C_k = \bigcap_{j=1}^{k} C_j$$

*We refer to the intersection* $\bigcap_{j=1}^{\infty} C_j$ *as a* ***countable intersection***

# Set theory (cont.)

## Example

Define the sets

- $C_1 = \{(0,0),(0,1),(1,1)\}$
- $C_2 = \{(1,1),(1,2),(2,1)\}$

Then,

$$C_1 \cap C_2 = \{(1,1)\}$$

■

## Example

Define the sets

- $C_1 = \{(x,y) : 0 \leq x + y \leq 1\}$
- $C_2 = \{(x,y) : 1 < x + y\}$

Then,

$$C_1 \cap C_2 = \emptyset$$

# Set theory (cont.)

## Example

For every $C$, $C \cap C = C$ and $C \cap \emptyset = \emptyset$

■

## Example

Let

$$C_k = \left\{x : 0 < x < \frac{1}{k}\right\}, \quad k = 1, 2, 3, \cdots$$

Then,

$$\bigcap_{k=1}^{\infty} C_k = \emptyset$$

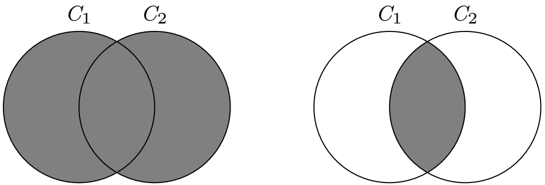
■

# Set theory (cont.)

**Example**

Let $C_1$ and $C_2$ represent the point sets enclosed by two intersecting balls

Then, sets $C_1 \cup C_2$ and $C_1 \cap C_2$ can be represented by the shaded regions




⇝ **Venn diagrams**

# Set theory (cont.)

## Definition

*In some situations, the entirety of all elements of interest can be described*

*The set of all elements of interest is the* ***space***

- *Spaces are denoted by calligraphic letters*
- *(such as $\mathcal{C}, \mathcal{D}, \cdots$)*

# Set theory (cont.)

## Example

### Multiple tosses of a coin

Tossing a coin four times, let $x$ be the number of heads

- Number $x$ is one of the numbers $0, 1, 2, 3, 4$

Then, the space is set $\mathcal{C} = \{0, 1, 2, 3, 4\}$

## Example

### Rectangles

Consider all possible rectangles of base $x$ and height $y$

- Both $x$ and $y$ must be positive

Then, the space is set $\mathcal{C} = \{(x, y) : x, y > 0\}$

# Set theory (cont.)

## Definition

*Let $\mathcal{C}$ be a space and let $C$ be a subset of set $\mathcal{C}$*

*The set of all elements of $\mathcal{C}$ that do not belong to $C$*

⇝ *The **complement** of $C$ ('with respect to $\mathcal{C}$')*

- *The complement of $C$ is indicated by $C^c$*

Special case, $\mathcal{C}^c = \emptyset$

# Set theory (cont.)

**Example**

Let $\mathcal{C} = \{0, 1, 2, 3, 4\}$ and let $C = \{0, 1\}$

The complement of $C$ (wrt $\mathcal{C}$) is $C^c = \{2, 3, 4\}$

■

# Set theory (cont.)

## Example

Let $C \subset \mathcal{C}$

We have that,

1. $C \cup C^c = \mathcal{C}$
2. $C \cap C^c = \emptyset$
3. $C \cup \mathcal{C} = \mathcal{C}$
4. $C \cap \mathcal{C} = C$
5. $(C^c)^c = C$

■

# Set theory (cont.)

## Example

**DeMorgan's laws**

Let $\mathcal{C}$ be a space and let $C_i \subset \mathcal{C}$, with $i = 1, 2$

Then,

- $(C_1 \cap C_2)^c = C_1^c \cup C_2^c$
- $(C_1 \cup C_2)^c = C_1^c \cap C_2^c$

■($\star$)

# Set theory (cont.)

Most of the functions used in calculus map real numbers into real numbers

- Our concern is mostly with functions that map sets into real numbers

Such functions are functions of a set

- **Set functions**

# Set theory (cont.)

## Example

Let $C$ be a set in a mono-dimensional space

Let $Q(C)$ be equal to the number of points in $C$ that are positive integers

Then, $Q(C)$ is a function of set $C$

- If $C = \{x : 0 < x < 5\}$, then $Q(C) = 4$
- If $C = \{-2, -1\}$, then $Q(C) = 0$
- If $C = \{x : -\infty < x < 6\}$, then $Q(C) = 5$

■

# Set theory (cont.)

## Example

Let $C$ be a set in a bi-dimensional space

Let $Q(C)$ be the area of $C$, if $C$ has a finite one

- Let $Q(C)$ remain unset, otherwise

Then, $Q(C)$ is a function of set $C$

- If $C = \{(x, y) : x^2 + y^2 \leq 1\}$, then $Q(C) = \pi$
- If $C = \{(0, 0), (1, 1), (0, 1)\}$, then $Q(C) = 0$
- If $C = \{(x, y) : 0 \leq x, 0 \leq y, x + y \leq 1\}$, then $Q(C) = 1/2$

■

# Set theory (cont.)

## Example

Let $C$ be a set in a tri-dimensional space

Let $Q(C)$ be the volume of $C$, if $C$ has a finite one

- Let $Q(C)$ remain unset, otherwise

Then, $Q(C)$ is a function of set $C$

- If $C = \{(x, y, z) : 0 \leq x \leq 2, 0 \leq y \leq 1, 0 \leq z \leq 3\}$, then $Q(C) = 6$
- If $C = \{(x, y, z) : x^2 + y^2 + z^2 \geq 1\}$, then $Q(C)$ is undefined

# Set theory (cont.)

We introduce the notation

The symbol

$$\int_C f(x)\mathrm{d}x$$

indicates the ordinary, Reimannian, integral of $f(x)$

- over the mono-dimensional set $C$

The symbol

$$\int\int_C g(x,y)\mathrm{d}x\mathrm{d}y$$

indicates the ordinary, Reimannian, integral of $g(x,y)$

- over the bi-dimensional set $C$

# Set theory (cont.)

**Remark**

Ordinary integrals have a bad habit: At times, they fail to exist

- Sets $C$ and functions $f(x)$ and $g(x, y)$ must be chosen with care

# Set theory (cont.)

The symbol

- $$\sum_C f(x)$$

  indicates the sum extended over all points $x \in C$

The symbol

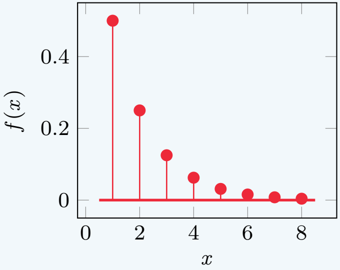
- $$\sum_C \sum g(x, y)$$

  indicates the sum extended over all points $(x, y) \in C$

# Set theory (cont.)

## Example

Let $C$ be a set in a mono-dimensional space and let $Q(C) = \sum_C f(x)$




$$f(x) = \begin{cases} \left(\frac{1}{2}\right)^x, & x = 1, 2, 3, \dots \\ 0, & \text{elsewhere} \end{cases}$$

If $C = \{x : 0 \leq x \leq 3\}$, then

$$Q(C) = \frac{1}{2} + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^3 = 7/8$$

# Set theory (cont.)

**Remark**

The simplest of all series of non-negative terms is the **geometric series**

If $0 \leq x < 1$, then

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$$

If $x \geq 1$, the series diverges

If $x \neq 1$,

$$s_n = \sum_{k=0}^{n} x^k = \frac{1 - x^{n+1}}{1-x}$$

Result follows if we let $n \to \infty$

For $x = 1$,

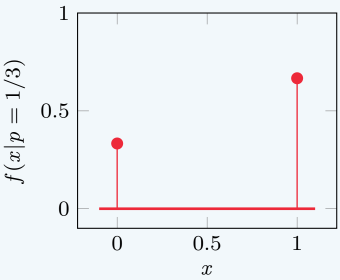
$$1 + 1 + 1 + \cdots$$

which diverges

# Set theory (cont.)

## Example

Let $C$ be a set in a mono-dimensional space and let $Q(C) = \sum_C f(x)$




$$f(x) = \begin{cases} p^x(1-p)^{1-x}, & x = 0, 1 \\ 0, & \text{elsewhere} \end{cases}$$

If $C = \{0\}$, then

$$Q(C) = \sum_{x=0}^{0} p^x(1-p)^{1-x} = 1 - p$$

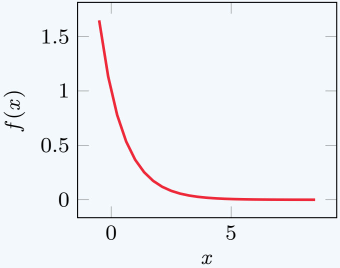
If $C = \{x : 1 \leq x \leq 2\}$, then

$$Q(C) = f(1|p) = p$$

# Set theory (cont.)

## Example

Let $C$ be a mono-dimensional set and let $Q(C) = \int_c f(x) \mathrm{d}x$




$f(x) = e^{-x}$

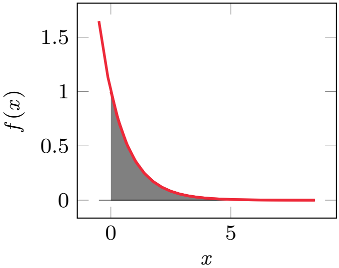
If $C = \{x : 0 \leq x \leq \infty\}$, then

$$Q(C) = \int_0^\infty e^{-x} \mathrm{d}x = 1$$

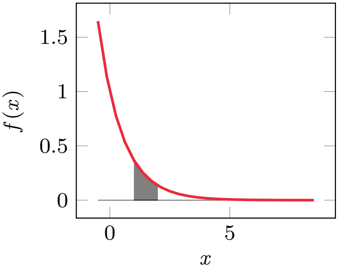
If $C = \{x : 1 \leq x \leq 2\}$, then

$$Q(C) = \int_1^2 e^{-x} \mathrm{d}x = e^{-1} - e^{-2}$$

# Set theory (cont.)



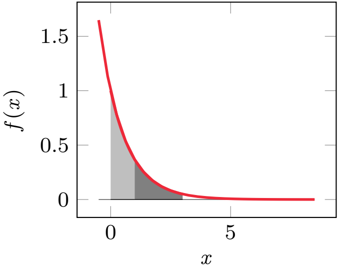

(a) $C = \{x : 0 \leq x \leq \infty\}$




(b) $C = \{x : 1 \leq x \leq 2\}$

# Set theory (cont.)

If $C_1 = \{x : 0 \leq x \leq 1\}$ and $C_2 = \{x : 1 < x \leq 3\}$, then




$$Q(C_1 \cup C_2) = \int_0^3 e^{-x} \mathrm{d}x = \int_0^1 e^{-x} \mathrm{d}x + \int_1^3 e^{-x} \mathrm{d}x = Q(C_1) + Q(C_2)$$

# Set theory (cont.)

## Example

Let $C$ be a set in a $n$-dimensional space and let

$$Q(C) = \int \cdots \int_C (1) \mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_n$$

If $C = \{(x_1, x_2, \dots, x_n) : 0 \leq x_1 \leq x_2 \leq \cdots \leq x_n \leq 1\}$, then

$$\begin{aligned} Q(C) &= \int_0^1 \int_0^{x_n} \cdots \int_0^{x_3} \int_0^{x_2} \mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_{n-1} \mathrm{d}x_n \\ &= \frac{1}{n!}, \quad \text{with } n! = n(n-1) \cdots 3 \cdot 2 \cdot 1 \end{aligned}$$

■

# Probability set functions

## Probability theory

# Probability set functions

Given an experiment, let $\mathcal{C}$ be the sample space (all possible outcomes)

- We want to assign probabilities to events (subsets of $\mathcal{C}$)

What our collection of events should be?

## Example

- If $\mathcal{C}$ is a finite set, we could pick the set of all subsets of $\mathcal{C}$

# Probability set functions (cont.)

We pick our collection of events in such a way that

1. It is sufficiently rich to incorporate all possible events (of interest)
2. It is closed under complements and countable unions of such events

The collection is, then, closed under countable intersections (DeMorgan)

A set of events thus defined is said to be a **Borel $\sigma$-field** of subsets

- To indicate it, we use symbol $\mathcal{B}$

## Example

Let $\mathcal{C} = \{\texttt{A}, \texttt{B}, \texttt{C}, \texttt{D}\}$, set $\mathcal{B} = \{\emptyset, \{\texttt{A}, \texttt{B}\}, \{\texttt{C}, \texttt{D}\}, \{\texttt{A}, \texttt{B}, \texttt{C}, \texttt{D}\}\}$ is a possible $\sigma$-field

- It contains the empty set
- It is closed under complements
- It is closed under countable unions of its elements

$\rightsquigarrow$ It is closed under countable intersections of its elements

The pair $(\mathcal{C}, \mathcal{B})$ is called a measurable space

# Probability set functions (cont.)

We have a sample space $\mathcal{C}$ and we have a collection of events $\mathcal{B}$

We can define the last component of a probability space

$\rightsquigarrow$ The probability set function

To push in the definition, we use relative frequencies

# Probability set functions (cont.)

The definition of probability consists of three axioms

- Motivation from three properties of relative frequency

## Remark

Let $\mathcal{C}$ be a sample space and let $C \subset \mathcal{C}$

- We repeat the experiment $N$ times
- The relative frequency of $C$ is $f_C = \#\{C\}/N$

$\#\{C\}$ is the number of times $C$ occurred in $N$ repetitions

$\rightsquigarrow$ $f_C \geq 0$ and $f_{\mathcal{C}} = 1$

- The first two properties

Let $C_1$ and $C_2$ be two disjoint events then

$\rightsquigarrow$ $f_{C_1 \cup C_2} = f_{C_1} + f_{C_2}$

- The third property

The three properties of relative frequencies form the axioms of probability

- The third axiom is given in terms of countable unions

# Probability set functions (cont.)

## Definition 3.1

***Probability***

*Let $\mathcal{C}$ be a sample space and let $\mathcal{B}$ be the set of events*

*Let $P$ be a real-valued function defined on $\mathcal{B}$*

*$P$ is a* ***probability set function****, if $P$ satisfies the three conditions*

1. $P(C) \geq 0$, *for all* $C \in \mathcal{B}$
2. $P(\mathcal{C}) = 1$
3. *If* $\{C_n\}$ *is a sequence of events in* $\mathcal{B}$ *and* $C_m \cap C_n = \emptyset$ *for all* $m \neq n$,

$$\rightsquigarrow P\Big(\bigcup_{n=1}^{\infty} C_n\Big) = \sum_{n=1}^{\infty} P(C_n)$$

# Probability set functions (cont.)

A collection of events whose elements are pair-wise disjoint (third condition)

$\rightsquigarrow$ **mutually exclusive** collection

A collection is **exhaustive** if the union of its events is the sample space

$$\rightsquigarrow \sum_{n=1}^{\infty} P(C_n) = 1$$

Mutually exclusive and exhaustive sets of events form a **partition** of $\mathcal{C}$

# Probability set functions (cont.)

A probability set function informs on how probability is distributed

- Over the set of events $\mathcal{B}$

$\rightsquigarrow$ A distribution of probability

We refer to $P$ as a probability (set) function

# Probability set functions (cont.)

The next theorems give some more properties of a probability set function

In the statement of each of these theorems, $P(C)$ is taken to be a probability set function defined on the collection of events $\mathcal{B}$ of some sample space $\mathcal{C}$

# Probability set function (cont.)

## Theorem 3.1

*For each event $C \in \mathcal{B}$, $P(C) = 1 - P(C^c)$*

**Proof**

We have $\mathcal{C} = C \cup C^c$ and $C \cap C^c = \emptyset$

From the second ($P(\mathcal{C}) = 1$) and the third ($P(\bigcup_{n=1}^{\infty} C_n) = \sum_{n=1}^{\infty} P(C_n)$, for disjoint sets) condition of the probability definition, it follows that

$$1 = P(C) + P(C^c)$$

■

# Probability set functions (cont.)

## Theorem 3.2

*The probability of the null set is zero*

$$P(\emptyset) = 0$$

**Proof**

From Theorem 3.1,

$$P(C) = 1 - P(C^c)$$

Take $C = \emptyset$ so that $C^c = \mathcal{C}$, thus

$$P(\emptyset) = 1 - P(\mathcal{C}) = 1 - 1 = 0$$

■

# Probability set functions (cont.)

## Theorem 3.3

*If $C_1$ and $C_2$ are events such that $C_1 \subset C_2$, then $P(C_1) \leq P(C_2)$*

**Proof**

$C_2 = C_1 \cup (C_1^c \cap C_2)$ and $C_1 \cap (C_1^c \cap C_2) = \emptyset$

From Definition 3.1,

$$P(C_2) = P(C_1) + P(C_1^c \cap C_2)$$

From the first condition ($P(C) \geq 0$, for all $C \in \mathcal{B}$) in Definition 3.1,

$$P(C_1^c \cap C_2) \geq 0$$

Hence, $P(C_2) \geq P(C_1)$

■

# Probability set functions (cont.)

## Theorem 3.4

*For each $C \in \mathcal{B}$, $0 \leq P(C) \leq 1$*

**Proof**

Since $\emptyset \subset C \subset \mathcal{C}$, by Theorem 3.3, we have that

$$P(\emptyset) \leq P(C) \leq P(\mathcal{C}) \quad \text{or} \quad 0 \leq P(C) \leq 1$$

■

# Probability set functions (cont.)

## Theorem 3.5

*If $C_1$ and $C_2$ are events in $\mathcal{C}$, then*

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

**Proof**

Sets $C_1 \cup C_2$ and $C_2$ can be represented by the union of non-relevant sets

$$C_1 \cup C_2 = C_1 \cup (C_1^c \cap C_2) \quad \text{and} \quad C_2 = (C_1 \cap C_2) \cup (C_1^c \cap C_2)$$

From the third condition ($P(\bigcup_{n=1}^{\infty} C_n) = \sum_{n=1}^{\infty} P(C_m)$, for disjoint sets)

$$P(C_1 \cup C_2) = P(C_1) + P(C_1^c \cap C_2) \quad \text{and} \quad P(C_2) = P(C_1 \cap C_2) + P(C_1^c \cap C_2)$$

Second equation is solved for $P(C_1^c \cap C_2)$, result plugged into the first one

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

■

# Probability set functions (cont.)

## Remark

**Inclusion-exclusion formula**

Let

- $p_1 = P(C_1) + P(C_2) + P(C_3)$
- $p_2 = P(C_1 \cap C_2) + P(C_1 \cap C_3) + P(C_2 \cap C_3)$
- $p_3 = P(C_1 \cap C_2 \cap C_3)$

Then, it can be shown that

$$P(C_1 \cup C_2 \cup C_3) = p_1 - p_2 + p_3,$$

This fact can be generalised: The **inclusion-exclusion formula**

$$P(C_1 \cup C_2 \cup \cdots \cup C_k) = p_1 - p_2 + p_3 - \cdots + (-1)^{k+1} p_k,$$

$p_i$ is the sum of probabilities of all possible intersections comprising $i$ sets

For $k = 3$, we have $p_1 \geq p_2 \geq p_3$ and more generally $p_1 \geq p_2 \geq \cdots \geq p_k$

# Probability set functions (cont.)

We shall show (Theorem 3.6) that

- $p_1 = P(C_1) + P(C_2) + \cdots + P(C_k) \geq P(C_1 \cup C_2 \cup \cdots \cup C_k)$

⇝ This expression is know as **Boole's inequality**

For $k = 2$, we get $1 \geq P(C_1 \cup C_2) = P(C_1) + P(C_2) - P(C_1 \cap C_2)$

This gives us

$$P(C_1 \cap C2) \geq P(C_1) + P(C_2) - 1 \quad (1)$$

⇝ This expression is known as **Bonferroni's inequality**

# Probability set functions (cont.)

The inclusion-exclusion formula provides other useful inequalities

$$p_1 \geq P(C_1 \cup C_2 \cup \cdots \cup C_k) \geq p_1 - p_2$$

and

$$p_1 - p_2 + p_3 \geq P(C_1 \cup C_2 \cup \cdots \cup C_k) \geq p_1 - p_2 + p_3 - p_4$$

■

# Probability set function (cont.)

## Example

Let the outcome of casting two coloured dice be an ordered pair

- (Number of spots on red die, Number of spots on white die)

The sample space is set $\mathcal{C} = \{(1,1), \cdots, (1,6), (2,1), \cdots, (2,6), \cdots, (6,6)\}$

Let the probability function assign probability 1/36 to the (36) points in $\mathcal{C}$

- (i.e., assume the dice are fair)

If

- $C_1 = \{(1,1), (2,1), (3,1), (4,1), (5,1)\}$
- $C_2 = \{(1,2), (2,2), (3,2)\}$

then,

- $P(C_1) = 5/36$
- $P(C_2) = 3/36$
- $P(C_1 \cup C_2) = 8/36$
- $P(C_1 \cap C_2) = 0$

# Probability set functions (cont.)

## Example

Consider two coins, the outcome is the ordered pair

$$\{(\texttt{Face on coin one}, \texttt{Face on coin two})\}$$

The sample space is set $\mathcal{C} = \{(\texttt{H}, \texttt{H}), (\texttt{H}, \texttt{T}), (\texttt{T}, \texttt{H}), (\texttt{T}, \texttt{T})\}$

A probability set function assign a probability of $1/4$ to each element of $\mathcal{C}$

Let

- $C_1 = \{(\texttt{H}, \texttt{H}), (\texttt{H}, \texttt{T})\}$, $P(C_1) = 1/2$
- $C_2 = \{(\texttt{H}, \texttt{H}), (\texttt{T}, \texttt{H})\}$, $P(C_2) = 1/2$

Then,

$$P(C_1 \cap C_2) = 1/4$$
$$P(C_1 \cup C_2) = 1/2 + 1/2 - 1/4 = 3/4 \text{ (Theorem 3.5)}$$

■

# Probability set functions (cont.)

## Example

**Equiprobable case**

Let $\mathcal{C}$ be partitioned into $k$ mutually disjoint subsets $C_1, C_2, \dots, C_k$

- Let the union of subsets $C_1, C_2, \dots, C_k$ be the sample space $\mathcal{C}$

Events $C_1, C_2, \dots, C_k$ are **mutually exclusive** and **exhaustive**

Suppose the random experiment is such that we can assume that each of the mutually exclusive and exhaustive events $C_i$ has the same probability

- It must be that $P(C_i) = 1/k$, with $i = 1, 2, \dots, k$
- Events $C_1, C_2, \dots, C_k$ are **equally probable**

# Probability set functions (cont.)

Let (event) $E$ be the union of $r$ of such mutually exclusive events

$$E = C_1 \cup C_2 \cup \cdots \cup C_r, \quad \text{with } r < k$$

Then,

$$P(E) = P(C_1) + P(C_2) + \cdots + P(r) = r/k$$

The experiment can terminate in $k$ number of ways (for such partition of $\mathcal{C}$)

⇝ The number of ways favourable to event $E$, $r$

In this sense, $P(E)$ is equal to the number of ways $r$ favourable to event $E$ divided by the total number of ways $k$ in which the experiment can conclude

# Probability set functions (cont.)

Setting probability $r/k$ to event $E$ assumes that each of the mutually exclusive and exhaustive events $C_1, C_2, \ldots, C_k$ has the same probability $1/k$

- The assumption of equiprobable events has become part of the model
- In some applications, this assumption is not realistic

■

# Probability set functions (cont.)

## Remark

**Counting rules**

Suppose we have two experiments

- The first experiment can result in $m$ outcomes
- The second one can result in $n$ outcomes

The composite experiment (1-st experiment followed by 2-nd experiment)

- $m \cdot n$ outcomes, represented by $m \cdot n$ ordered pairs

The **multiplication rule**: It extends to more than two experiments

# Probability set functions (cont.)

Let $A$ be a set with $n$ elements

Suppose we are interested in $k$-tuples whose elements are elements of $A$

There are $\underbrace{n \cdot n \cdots n}_{k \text{ factors}} = n^k$ such $k$-tuples

# Probability set functions (cont.)

Suppose $k \leq n$

We are interested in $k$-tuple whose members are distinct elements of $A$

There are,

- $n$ elements from which to choose for the first member
- $n - 1$ elements from which to choose for the second member
- $\cdots$
- $n - (k - 1)$ elements from which to choose for the $k$-th member

# Probability set functions (cont.)

By multiplication rule, there are $\underbrace{n \cdot (n-1) \cdots [n-(k-1)]}_{k \text{ factors}}$ such $k$-tuples

- Each such $k$-tuple is called a **permutation**

We let symbol $P_k^n$ to indicate the $k$ permutations, picked from $n$ elements

$$P_k^n = n(n-1)\cdots[n-(k-1)] = n(n-1)\cdots[n-(k-1)] \cdot \frac{(n-k)!}{(n-k)!} = \frac{n!}{(n-k)!} \quad (2)$$

# Probability set functions (cont.)

Suppose that order is no longer important, we count the number of subsets of $k$ elements taken from $A$, instead of counting the number of permutations

We let symbol $\binom{n}{k}$ indicate the total number of such subsets

Consider a subset of $k$ elements from $A$

By multiplication rule, it gives us $P_k^k = \underbrace{k(k-1)\cdots 1}_{k \text{ factors}}$ permutations

- All such permutations are different from the permutations obtained from other subsets of $k$ elements from $A$
- Each permutation of $k$ different elements from $A$ must be obtained from one of these subsets, of which there are $\binom{n}{k}$

Hence, we have

$$P_k^n = \binom{n}{k} k! \qquad \rightsquigarrow \qquad \binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{3}$$

# Probability set functions (cont.)

## Remark

Instead of subsets, we often prefer the terminology **combinations**

We say there are $\binom{n}{k}$ combinations of $k$ things picked from a set of $n$ things

## Remark

If the binomial is expanded, we get

$$(a+b)^n = (a+b)(a+b)\cdots(a+b) = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

We can select the $k$ factors from which to pick $a$ in $\binom{n}{k}$ ways

$\rightsquigarrow$ $\binom{n}{k}$ is called the **binomial coefficient**

# Probability set functions (cont.)

## Example

Let a card be picked at random from an ordinary deck (52 playing cards)

The sample space $\mathcal{C}$ is the union of $k = 52$ outcomes

- We assume each has equal probability 1/52
- (13 cards for each of the 4 suits)

If $E_1$ is the set of outcomes that are ♠, $P(E_1) = 13/52 = 1/4$

- 1/4 is the probability of drawing a ♠
- There are $r_1 = 13$ ♠s in the deck

If $E_2$ is the set of outcomes that are K, $P(E_2) = 4/52 = 1/13$

- 1/13 is the probability of drawing a K
- There are $r_2 = 4$ Ks in the deck

# Probability set functions (cont.)

Instead of picking a single card, suppose we pick five of them

- At random, without replacement
- Order is not important

This is a subset of 5 elements drawn from a set of 52 elements

- By Equation 3, there are $\binom{n=52}{k=5}$ possible draws

They should all be equally probable

- Each has probability $1/\binom{52}{5}$
- (well-shuffled deck)

# Probability set functions (cont.)

Let $E_3$ be the event of a flush (five cards of the same suit)

- There are $\binom{4}{1}$ suits to pick for a flush
- In each suit, there are $\binom{13}{5}$ possible hands

By the multiplication rule, the probability of a flush

$$P(E_3) = \binom{4}{1}\binom{13}{5} \Big/ \binom{52}{5} = \frac{4 \times 1,287}{2,598,960} \approx 0.002$$

# Probability set functions (cont.)

Consider the probability of getting three cards of the same kind (event $E_4$)

1. Select the kind, $\binom{13}{1}$ ways
2. Select the three, $\binom{4}{3}$ ways
3. Select the other two kinds, $\binom{12}{2}$ ways
4. Select one card from each of such last two kinds, $\binom{4}{1}\binom{4}{1}$ ways

The probability of exactly three cards of the same kind

$$P(E_4) = \binom{13}{1}\binom{4}{3}\binom{12}{2}\binom{4}{1}^2 \Big/ \binom{52}{5} \approx 0.02$$

# Probability set functions (cont.)

Suppose $E_5$ is the set of outcomes in which three cards are K and two are Q

1. Pick the K's in $\binom{4}{3}$ ways
2. Pick the Q's in $\binom{4}{2}$ ways

The probability of $E_5$ is

$$P(E_5) = \binom{4}{3}\binom{4}{2} \Big/ \binom{52}{2} \approx 0.000009$$

■

# Probability set functions (cont.)

We have shown one way in which we can define a probability set function

- A set function that satisfies Definition 3.1

Suppose the sample space $\mathcal{C}$ consists of $k$ distinct points

- (A mono-dimensional space, just for this discussion)

If the random experiment that ends in one of those $k$ points is such that it makes sense to assume that these points are equally probable, then

- ⇝ We can assign $1/k$ to each point
- ⇝ For $C \subset \mathcal{C}$, we let

$$P(C) = \frac{\text{number of points in } C}{k} = \sum_{x \in C} f(x), \quad \text{where } f(x) = 1/k, \quad x \in \mathcal{C}$$

# Probability set functions (cont.)

## Example

In the cast of a die, we take $\mathcal{C} = \{1, 2, 3, 4, 5, 6\}$ and $f(x) = 1/6, x \in \mathcal{C}$

- (If we believe that the die is unbiased)

'Unbiased': It warns us on the possibility that all points (six) might NOT be equally probable in all these cases (and loaded dice do, in fact, exist)

Suppose that a die has been biased so that the relative frequencies of the number in $\mathcal{C}$ appear to stabilise proportional to the number of up-face spots

⇝ We may assign $f(x) = x/21$ with $x \in \mathcal{C}$, and

$$P(C) = \sum_x f(x) \quad \text{[Definition (3.1)]}$$

Then, for $C = \{1, 2, 3\}$

$$P(C) = \sum_{x=1}^{3} f(x) = \frac{1}{21} + \frac{2}{21} + \frac{3}{21} = \frac{6}{21} = 2/7$$

■

# Probability set functions (cont.)

A sequence $\{C_n\}$ of events is non-decreasing if $C_n \subset C_{n+1}$, for all $n$

$\rightsquigarrow$ By definition, we have $\lim_{n\uparrow\infty} C_n = \bigcup_{n=1}^{\infty} C_n$

Consider $\lim_{n\uparrow\infty} P(C_n)$, can we swap limit and $P$?

Yes, the result holds for decreasing sequences of events too

# Probability set functions (cont.)

## Theorem

***Continuity theorem of probability***

*Let* $\{C_n\}$ *be a non-decreasing sequence of events, then*

$$\lim_{n \uparrow \infty} P(C_n) = P(\lim_{n \uparrow C_n} C_n) = P\Big(\bigcup_{n=1}^{\infty} C_n\Big) \tag{4}$$

*Let* $\{C_n\}$ *be a decreasing sequence of events, then*

$$\lim_{n \uparrow \infty} P(C_n) = P(\lim_{n \uparrow \infty} C_n) = P\Big(\bigcap_{n=1}^{\infty} C_n\Big) \tag{5}$$

# Probability set functions (cont.)

## Proof

We prove Equation (4) [Equation (5), ($\star$)]

Define the sets (rings) as $R_1 = C_1$ and, for $n > 1$, $R_n = C_n \cap C_{n-1}^c$

It follows that

- $\bigcup_{n=1}^{\infty} C_n = \bigcup_{n=1}^{\infty} R_n$
- $R_m \cap R_n = \emptyset$, for $m \neq n$

Also, $P(R_n) = P(C_n) - P(C_{n-1})$

Because of the third axiom of probability,

$$P\Big[\lim_{n\uparrow\infty} C_n\Big] = P\Big(\bigcup_{n=1}^{\infty} C_n\Big) = P\Big(\bigcup_{n=1}^{\infty} R_n\Big) = \sum_{n=1}^{\infty} P(R_n) = \lim_{n\uparrow\infty} \sum_{j=1}^{n} P(R_n)$$
$$= \lim_{n\uparrow\infty} \Big\{P(C_1) + \sum_{j=2}^{n} \Big[P(C_j) - P(C_{j-1})\Big]\Big\} = \lim_{n\uparrow\infty} P(C_n) \quad (6)$$

■

# Probability set functions (cont.)

## Theorem 3.6

***Boole's inequality***

*Let $\{C_n\}$ be an arbitrary sequence of events*

*Then,*

$$P\Big(\bigcup_{n=1}^{\infty} C_n\Big) \leq \sum_{n=1}^{\infty} P(C_n) \tag{7}$$

**Proof**

Let $D_n = \bigcup_{i=1}^{n} C_i$, then $\{D_n\}$ is an increasing sequence of events

- It goes up to $\bigcup_{n=1}^{\infty} C_n$

For all $j$, we have that $D_j = D_{j-1} \cup C_j$

Hence, by Theorem (3.5)

$$P(D_j) \leq P(D_{j-1}) + P(C_j) \longrightarrow P(D_j) - P(D_{j-1}) \leq P(C_j)$$

# Probability set function (cont.)

In this case, the $D_j$s replace the $C_j$s in the expression in Equation (6)

Thus, using the inequality above in Equation (6) and $P(C_1) = P(D_1)$

$$P\Big(\bigcup_{n=1}^{\infty} C_n\Big) = P\Big(\bigcup_{n=1}^{\infty} D_n\Big) = \lim_{n \uparrow \infty} \Big\{ P(D_1) + \sum_{j=2}^{n} \Big[ P(D_j) - P(D_{j-1}) \Big] \Big\}$$

$$\leq \lim_{n \uparrow \infty} \sum_{j=1}^{n} P(C_j) = \sum_{n=1}^{\infty} P(C_n)$$

■

# Conditional probability

## Probability theory

# Conditional probability

In some random experiments, our interest is only in some of the outcomes

- The elements of a subset $C_1$ of sample space $\mathcal{C}$

$\rightsquigarrow$ This means that our sample space is subset $C_1$

With $C_1$ as new sample space, we shall define a probability set function

# Conditional probability (cont.)

Let the probability set function $P(C)$ be defined on the sample space $\mathcal{C}$

- Let $C_1$ be a subset of $\mathcal{C}$ such that $P(C_1) > 0$

We consider only outcomes of the experiment that are elements of $C_1$

- We take $C_1$ to be a sample space

Let $C_2$ be another subset of $\mathcal{C}$

# Conditional probability (cont.)

How to define the probability of event $C_2$, relative to sample space $C_1$?

- This probability is (will be) the **conditional probability** of event $C_2$, relative to the hypothesis of event $C_1$ (once defined)
- More briefly, the conditional probability of $C_2$, given $C_1$

Such a conditional probability is indicated by the symbol $P(C_2|C_1)$

# Conditional probability (cont.)

Relative to sample space $C_1$, how to define the probability of event $C_2$?

As $C_1$ is the sample space, the elements of $C_2$ that are of interest are those (if any) that are also elements of $C_1$ (that is, the elements of $C_1 \cap C_2$)

It is beneficial to define the symbol $P(C_2|C_1)$ so that

- $P(C_1|C_1) = 1$
- $P(C_2|C_1) = P(C_1 \cap C_2|C_1)$

# Conditional probability (cont.)

The ratio of probabilities of events $C_1 \cap C_2$ and $C_1$, relative to space $C_1$ and the ratio of probabilities of these events relative to space $\mathcal{C}$, should be

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

# Conditional probability (cont.)

- $P(C_1|C_1) = 1$
- $P(C_2|C_1) = P(C_1 \cap C_2|C_1)$
- $\dfrac{P(C_1 \cap C_2|C_1)}{P(C_1|C_1)} = \dfrac{P(C_1 \cap C_2)}{P(C_1)}$

These three conditions imply the relation

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

Such a relation is a fine definition of **conditional probability** of event $C_2$

- Given event $C_1$ and provided that $P(C_1) > 0$

# Conditional probability (cont.)

Moreover

1. $P(C_2|C_1) \geq 0$
2. Provided that $C_2, C_3, \dots$ are mutually exclusive events

$$P\Big(\bigcup_{j=2}^{\infty} C_j|C_1\Big) = \sum_{j=2}^{\infty} P(C_j|C_1)$$

3. $P(C_1|C_1) = 1$

These are the conditions that a probability set function must satisfy

⇝ $P(C_2|C_1)$ is a probability function, defined for subsets of $C_1$

# Conditional probability (cont.)

$P(C_2|C_1)$ is the conditional probability function, relative to hypothesis $C_1$

- Or, the conditional probability set function, given $C_1$

This conditional probability function, given $C_1$, is defined when $P(C_1) > 0$

# Conditional probability (cont.)

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

From the definition of the conditional probability set function,

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

This relation is the **multiplication rule** for probabilities

The nature of the experiment, might suggests possible assumptions

- ⇝ How $P(C_1)$ and $P(C_2|C_1)$ can be assigned

Then, $P(C_1 \cap C_2)$ can be computed under these assumptions

# Conditional probability (cont.)

## Example

A box contains eight stones, three of which are red and five are blue

- Two stones are randomly drawn successively, no replacement

What is the probability that

1. the first draw results in a red stone (event $C_1$)?
2. the second draw results in a blue stone (event $C_2$)?

It is reasonable to assign the following probabilities

- $P(C_1) = 3/8$
- $P(C_2|C_1) = 5/7$

Given these assignments, we have

$$P(C_1 \cap C_2) = \underbrace{(3/8)}_{P(C_1)} \underbrace{(5/7)}_{P(C_2|C_1)} = 15/56 = 0.2679$$

■

# Conditional probability (cont.)

The multiplication rule can be extended to three or more events

With three events, by the multiplication rule for two events

$$\begin{aligned} P(C_1 \cap C_2 \cap C_3) &= P\big[(C_1 \cap C_2) \cap C3)\big] \\ &= P(C_1 \cap C_2) P(C_3 | C_1 \cap C_2) \end{aligned}$$

But, $P(C_1 \cap C_2) = P(C_1)P(C_2|C_1)$, hence $P(C_1 \cap C_2) > 0$

$$P(C_1 \cap C_2 \cap C_3) = P(C_1)P(C_2|C_1)P(C_3|C_1 \cap C_2)$$

■

## Remark

The procedure can be used to extend the multiplication rule to 4+ events

- The expression for $k$ events can be found by induction

# Conditional probability (cont.)

## Example

Four playing cards drawn from an ordinary deck successively, at random

- No replacement

The probability of a ♠, a ♡, a ♢ and a ♣ in that order

$$\big(13/52\big)\big(13/51\big)\big(13/50\big)\big(13/49\big) = 0.0044$$

■

# Conditional probability (cont.)

Consider $k$ mutually exclusive and exhaustive events $C_1, C_2, \dots, C_k$

- $P(C_i) > 0$, for $i = 1, 2, \dots, k$

⇝ $C_1, C_2, \dots, C_k$ form a partition of $\mathcal{C}$

The events $C_1, C_2, \dots, C_k$ need not be equally likely

Let $C$ be some another event such that $P(C) > 0$

$C$ occurs with one, and only one, of events $C_1, C_2, \dots, C_k$

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

# Conditional probability (cont.)

Given that $C \cap C_i, i = 1, 2, \dots, k$ are mutually exclusive

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

However,

$$P(C \cap C_i) = P(C_i)P(C|C_i), i = 1, 2, \dots, k$$

So,

$$\begin{aligned} P(C) &= P(C_1)P(C|C_1) + P(C_2)P(C|C_2) + \cdots + P(C_k)P(C|C_k) \\ &= \sum_{i=1}^{k} P(C_i)P(C|C_i) \end{aligned}$$

This result is known as **law of total probability**

# Conditional probability (cont.)

From the definition of conditional probability, by the law of total probability

$$P(C_j|C) = \frac{P(C \cap C_j)}{P(C)} = \frac{P(C_j)P(C|C_j)}{\sum_{i=1}^{k} P(C_i)P(C|C_i)} \quad (8)$$

This is known as the **Bayes' rule**

Get the conditional probability of $C_j$, given $C$, from the probabilities of $C_1, C_2, \ldots, C_k$ and the conditional probabilities of $C$, given $C_i$, $i = 1, 2, \ldots, k$

# Conditional probability (cont.)

## Example

- Box $C_1$ contains three red stones and seven blue stones
- Box $C_2$ contains eight red stones and two blue stones

All stones are identical in size, shape and weigth

Assume (reasonably?) that

1. the probability of selecting box $C_1$ is $P(C_1) = 1/3$
2. the probability of selecting box $C_2$ is $P(C_2) = 2/3$

A box is picked a random and then a stone is picked from it at random

- Suppose the stone is red
- We indicate this event by $C$

# Conditional probability (cont.)

By considering the boxes' content, we assign the conditional probabilities

1. $P(C|C_1) = 3/10$
2. $P(C|C_2) = 8/10$

The conditional probability of box $C_1$, given a red stone is picked

$$\begin{aligned} P(C_1|C) &= \frac{P(C_1)P(C|C_1)}{P(C_1)P(C|C_1) + P(C_2)P(C|C_2)} \\ &= \frac{(1/3)(3/10)}{(1/3)(3/10) + (2/3)(8/10)} \\ &\approx 0.16 \end{aligned}$$

$\rightsquigarrow$ $P(C_2|C) \approx 0.84$

# Conditional probability (cont.)

Probabilities $P(C_1) = 1/3$ and $P(C_2) = 2/3$ are called **prior probabilities**

- Assumed to be due to the random mechanism used to pick boxes

If a stone is picked and observed to be red, conditional probabilities $P(C_1|C) = 3/19$ and $P(C_2|C) = 16/19$ are then updated to **posterior probabilities**

# Conditional probability (cont.)

$C_2$ has a larger proportion of red stones (8/10) than $C_1$ (3/10), by intuition

- $P(C_2|C)$ should be larger than $P(C_2)$
- $P(C_1|C)$ should be smaller than $P(C_1)$

The chances of having picked box $C_2$ are higher once a red stone is observed

- (than before the stone is taken)

Bayes' rule provides a method of determining those probabilities

# Conditional probability (cont.)

## Example

Plants $C_1$, $C_2$ and $C_3$ produce 10%, 50% and 40% of the company's product

- Plant $C_1$ is small with only 1% of faulty products
- Plant $C_2$ and $C_3$ produce items that are 3% and 4% faulty

Products are shipped to central control where one piece is selected at random

- Then, it is found faulty (event $C$)

What is the (conditional) probability that the piece comes from plant $C_1$?

We can assign the prior probabilities of getting a(ny) piece from the plants

- $P(C_1) = 0.1$
- $P(C_2) = 0.5$
- $P(C_3) = 0.4$

# Conditional probability (cont.)

As for the conditional probabilities of faulty pieces, they are assigned as

- $P(C|C_1) = 0.01$
- $P(C|C_2) = 0.03$
- $P(C|C_3) = 0.04$

The posterior probability of $C_1$, given a faulty piece, is thus given by

$$P(C_1|C) = \frac{P(C_1 \cap C)}{P(C)} = \frac{(0.1)(0.01)}{(0.1)(0.01) + (0.5)(0.03) + (0.4)(0.04)} = 1/32$$

■

# Independence

## Probability theory

# Independence

The occurrence of event $C_1$ may not alter the probability of event $C_2$

⇝ $P(C_1|C_2) = P(C_2)$, for $P(C_1) > 0$

Such events $C_1$ and $C_2$ are said to be **independent**

The multiplication rule comes to be

$$P(C_1 \cap C_2) = P(C_1)P(C_2|C_1) = P(C_1)P(C_2) \tag{9}$$

This implies, when $P(C_2) > 0$, that

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

# Independence (cont.)

If $P(C_1) > 0$ and $P(C_2) > 0$, then independence is equivalent to

$$P(C_1 \cap C_2) = P(C_1)P(C_2) \tag{10}$$

What if $P(C_1) = 0$ and $P(C_2) = 0$?

- Either way, the RHS is zero
- The LHS is zero also because $C_1 \cap C_2 \subset C_1$ and $C_1 \cap C_2 \subset C_2$

# Independence (cont.)

**Definition**

*Let $C_1$ and $C_2$ be two events*

*$C_1$ and $C_2$ are said to be **independent** if*

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

Suppose $C_1$ and $C_2$ are independent events

Then,

- $C_1$ and $C_2^c$ are also independent
- $C_1^c$ and $C_2$ are also independent
- $C_1^c$ and $C_2^c$ are also independent

# Independence (cont.)

## Remark

Events that are independent are called stochastically independent

- Independent in a probabilistic sense
- Statistically independent

We use independent, without adjectives

■

# Independence (cont.)

Suppose that we have three events $C_1$, $C_2$ and $C_3$

They are **mutually independent** iff they are pair-wise independent

$$P(C_1 \cap C_2) = P(C_1)P(C_2)$$
$$P(C_1 \cap C_3) = P(C_1)P(C_3)$$
$$P(C_2 \cap C_3) = P(C_2)P(C_3)$$

and

$$P(C_1 \cap C_2 \cap C_3) = P(C_1)P(C_2)P(C_3)$$

# Independence (cont.)

In general, $n$ events $C_1, C_2, \ldots, C_n$ are **mutually independent** if and only if for every collection of $k$ of such events $(2 \leq k \leq n)$ the following holds

Let $d_1, d_2, \ldots, d_n$ be $k$ distinct integers from $1, 2, \ldots, n$, then

$$P(C_{d_1} \cap C_{d_2} \cap \cdots \cap C_{d_k}) = P(C_{d_1})P(C_{d_2}) \cdots P(C_{d_k})$$

## Remark

Specifically, if $C_1, C_2, \ldots, C_n$ are mutually independent, then

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

# Independence (cont.)

As with two sets, event combinations and complements are independent

- Events $C_1^c$ and $C_2 \cup C_3^c \cup C_4$ are independent
- Events $C_1 \cup C_2^c$, $C_3^c$, $C_4 \cap C_5^c$ are mutually independent

# Independence (cont.)

If no risk of misunderstanding, use 'independent' w/o adjective 'mutually'

- (also for more than two events)

**Remark**

**Pairwise independence does not imply mutual independence**

# Independence (cont.)

We may perform sequences of experiments so that the events associated with one of them are independent of the events associated with other experiments

- The events are referred to as outcomes of **independent experiments**
- The respective events are understood as independent

Thus we often refer to independent trials of some given random experiment

# Conditional probability and independence (cont.)

## Example

A coin is flipped independently a number of times

Let event $C_i$ be a head H on the $i$-th toss

- $\rightsquigarrow$ $C_i^c$ indicates tails T

Assume $C_i$ and $C_i^c$ as equally probable

$$P(C_i) = P(C_i^c) = 1/2$$

From independence, the probability for ordered sequences as HHTH

$$P(C_1 \cap C_2 \cap C_3^c \cap C_4) = P(C_1)P(C_2)P(C_3^c)P(C_4) = (1/2)^4 = 1/16$$

# Independence (cont.)

The probability of getting the first head on the third toss

$$P(C_1^c \cap C_2^c \cap C_3) = P(C_1^c)P(C_2^c)P(C_3) = (1/2)^3 = 1/8$$

The probability of observing at least one head on four tosses

$$\begin{aligned} P(C_1 \cup C_2 \cup C_3 \cup C_4) &= 1 - P\big[(C_1 \cup C_2 \cup C_3 \cup C_4)^c\big] \\ &= 1 - P(C_1^c \cap C_2^c \cap C_3^c \cap C_4^c) \\ &= 1 - (1/2)^4 = 15/16 \end{aligned}$$

■

# Independence (cont.)

## Example

A system such that if component $K_1$ fails, it is by-passed and component $K_2$ is used instead, if also component $K_2$ fails, then component $K_3$ is used

Suppose that the following probabilities are given

- Probability that $K_1$ fails is 0.01
- Probability that $K_2$ fails is 0.03
- Probability that $K_3$ fails is 0.10

Let failures be mutually independent events

The probability of system's failure as all components would have to fail is

$$(0.01)(0.03)(0.10) = 0.00003$$

The probability of not-failure

$$1 - 0.00003 = 0.99997$$

■