

# STAT 230

Notes: Syed Mustafa Raza Rizvi

June 1, 2020

These are my notes for the Probability course at the University of Waterloo (STAT 230).

They are pretty similar to the online lecture content

You will find that these aren't very useful as notes, in the sense that they are not significantly shorter than the content in the lecture content, they're really just a way for me to type down the content I am learning and absorb it.

Thanks to Prof. Dan Wolczuk for providing me with the macros to typeset this LaTeX document.

# Week 1

## 1 Introduction to Probability

### Probability

- A strong likelihood or chance of something
- The relative possibility an event will occur
- The ratio of the number of actual occurrences to the total number of possible occurrences

In this course we will consider so-called "random" experiments that have several possible outcomes and are repeatable

### Definitions of Probability

Let  $\mathcal{S}$  be the set of all possible distinct outcomes of a random experiment. Then the probability of an event, provided that all outcomes are equally likely, is

$$\frac{\text{Number of ways the event can occur}}{\text{Total number of outcomes in } \mathcal{S}}$$

Definition

Classical

Definition of  
Probability

The probability of an event in an experiment is the (limiting) proportion or fraction of times the event occurs in a very long (theoretically infinite) series of (independent) repetitions of the experiment

Definition

Relative

Frequency

Definition

The probability of an event is a 'best guess' by a person making the statement of the chances that the event will happen (e.g., 30% chance of rain)

Definition

Subjective

Probability

The classical definition and the relative frequency definition are **consistent** with one another if we are careful in constructing our model.

### Characteristics of a Random Experiment

- It should have more than one possible outcome
- We should be able to repeat the experiment under similar/identical conditions
- It *may* have equally likely outcomes

### Probability Models

- A sample space of all possible outcomes of a random experiment must be defined
- A set of events is defined. An event is a subset of the sample space, to which we can assign a probability
- A way of assigning probabilities, which are numbers between 0 and 1, to events is specified.

## 2 Random Experiments and Sample spaces

When we repeat the experiment under *controlled conditions*, (repetitions are called **trials** of the experiment) different outcomes may occur

Definition  
Random  
Experiment

### Properties of a Random experiment

- We should be able to repeat it
- Different outcomes may occur on different trials even if the conditions are the same
- Outcomes have probabilities associated with them

A sample space,  $\mathcal{S}$ , is the set of distinct outcomes for an experiment or process, with the property that in a single trial, one and only one of these outcomes occurs

Definition  
Sample

The sample space is a set and the outcomes in a sample space are called **sample points** or **points**

Space

A **discrete sample space**,  $\mathcal{S}$ , is one with a **finite** number of sample points or **countably many** sample points

Definition  
Discrete  
Sample  
Space

A set,  $\mathcal{S}$ , is countable if the elements can be put in a 1-1 correspondence with the positive integers

Definition  
Countable

E.g. A **countably infinite** set is the sample space of a coin tossing experiment that ends on the event that a Tails occurs. The set

$$\mathcal{S} = \{T, HT, HHT, HHHT, \dots\}$$

is countably infinite as we can form a 1-1 correspondence of each sample point with the positive integers, and since it is countable, it is discrete

Example of an **uncountably infinite** set is the sample space

$$\mathcal{S} = [0, \infty)$$

which has an infinite number of sample points, they **can not** be put into a 1-1 correspondence with the positive integers. The sample space  $\mathcal{S}$  is not discrete

### 3 Probability Models and Events

An **event**,  $A$ , defined on a discrete sample space,  $\mathcal{S}$ , is a subset of  $\mathcal{S}$ . i.e.  $A \subset \mathcal{S}$

Definition

Event

If the event  $A \subset \mathcal{S}$  consists of only one sample point this  $A$  is called a **simple event**

Definition

Simple Event

If the event  $A \subset \mathcal{S}$  consists of two or more sample points then  $A$  is called a **compound event**.  
 $A$  is said to **occur** on a trial of the experiment if one of the simple events in  $A$  occurs

Definition

Compound  
Event

Let  $\mathcal{S} = \{a_1, a_2, a_3, \dots\}$  be a discrete sample space

Definition

Let  $P(a_1), P(a_2), P(a_3), \dots$  be a set of numbers associated with the sample points  $a_1, a_2, a_3, \dots$  such that:

Probability  
Distribution

1)  $0 \leq P(a_i) \leq 1, i = 1, 2, \dots$

on  $\mathcal{S}$

2)  $\sum_{i=1} P(a_i) = 1$

Then  $P(a_i)$  is called a probability

The set  $\{P(a_i), i = 1, 2, \dots\}$  is called the **probability distribution on  $\mathcal{S}$**

The function  $P(*)$  has the sample space  $\mathcal{S}$  as its domain, the condition  $\sum P(a_i) = 1$  reflects the idea that when the process or experiment happens, one or other of the simple events  $a_i \in \mathcal{S}$  must occur. (The sum shows that it is certain the total probability of an outcome coming from the defined events in the sample space is 1, and hence 0 for anything outside the sample space).  
Events consisting of a single  $a_i$  in the sample space are referred to as **simple events**

Let  $\mathcal{S}$  be a discrete sample space and let  $A$  be an event defined on  $\mathcal{S}$ , i.e.  $A \subset \mathcal{S}$

Definition

Then  $P(A)$ , the probability of event  $A$ , is the sum of the probabilities corresponding to the sum of all the simple events that make up  $A$ ,  $P(A) = \sum_{a \in A} P(a)$

Probability  
of an Event

According to this definition  $P(A) = P(a_1) + P(a_2)$  would be the probability of a compound event  $A = \{a_1, a_2\}$

Note that for any event  $A, 0 \leq P(A) \leq 1$

A discrete sample space  $\mathcal{S} = \{a_1, a_2, \dots\}$  together with a probability distribution  $\{P(a_i), i = 1, 2, \dots\}$  is referred to as a **discrete probability model**

Definition

Discrete  
Probability  
Model

The odds in favour of an event  $A$  is the probability the event occurs divided by the probability it does not occur or  $\frac{P(A)}{1-P(A)}$ . The odds against the event is the reciprocal,  $\frac{1-P(A)}{P(A)}$

Definition

Odds

## 4 Counting Techniques

When all the simple events have the same probability, for any even  $A \subset \mathcal{S}$ ,

$$P(A) = \frac{\text{number of points in } A}{N(\text{number of points in } \mathcal{S})}$$

Definition

$P(A)$  for  
equi-  
probable  
outcomes

Suppose we can do job 1 in  $p$  ways and job 2 in  $q$  ways. Then we can do either job 1 **OR** job 2 (but not both), in a total of  $p + q$  ways

Definition

Addition  
Rule

Suppose we can do job 1 in  $p$  ways and, for each of these ways, we can do job 2 in  $q$  ways. Then we can do both job 1 **AND** job 2, in a total of  $pq$  distinct ways

Definition

Multiplication  
Rule

### Some useful combinatorial symbols

- $n^{(k)} = n(n-1)(n-2)\dots(n-k+1)$  (also written as  ${}^nP_r$ )  
= number of arrangements of  $n$  different elements taken  $k$  at a time
- $n! = n(n-1)\dots(2)(1)$   
= number of arrangements (permutations) of  $n$  different elements taken  $n$  at a time
- $n^k = n\dots n$   
= number of arrangements of  $n$  elements taken  $k$  at a time allowing repeats
- By definition  $0^0 = 1$ , and  $0! = 1$ , therefore  $0^{(0)} = 1$
- Note  $n^{(k)} = \frac{n!}{(n-k)!}$  when  $k \geq 0$  is an integer
- Note that  $n^{(k)}$  is also defined for a real number  $n$ , with  $k$  as a non-negative integer (e.g.  $e^{(3)} = e(e-1)(e-2)$ ,  $3^{(4)} = 3 \times 2 \times 1 \times 0$ )

Definition

$n$  choose  $k$

$$\binom{n}{k} = \frac{n^{(k)}}{k!} = \frac{n!}{(n-k)!k!} = {}^nC_k$$

If  $n$  is a positive integer and  $k$  is a non-negative integer such that  $k \leq n$  then  ${}^nC_k$  is the number of subsets (combinations) of  $k$  elements which may be selected from a set containing  $n$  elements

Definition

Multinomial  
Coefficients

$$\binom{n}{n_1, \dots, n_k} = \frac{n!}{n_1! \dots n_k!}$$

# Week 2

## Sets and Probability

Suppose  $\mathcal{S}$  is a sample space for an experiment

Let  $A, B, C, A_1, A_2, \dots$  be events defined on  $\mathcal{S}$

The union of  $A$  and  $B$  (written  $A \cup B$ ) is the set of all points which are in either  $A$  or  $B$  or both

Definition  
Union of  
Two Sets

The intersection of  $A$  and  $B$  (written  $A \cap B$ ) is the set of all points which are in both  $A$  and  $B$

Definition  
Intersection  
of Two Sets

The complement of  $A$  (written  $\bar{A}$ ) is the set of all points in  $\mathcal{S}$  which are not in  $A$

Definition  
Complement  
of a Set

The complement of the union is the intersection of the complements:

$$\overline{(A \cup B)} = \bar{A} \cap \bar{B}$$

De  
Morgan's  
Laws

The complement of the intersection is the union of the complements:

$$\overline{(A \cap B)} = \bar{A} \cup \bar{B}$$

Two events  $A$  and  $B$  are **mutually exclusive** if  $A \cap B = \emptyset$ , where  $\emptyset$  is the empty set

Definition

The events  $A_1, \dots, A_k$  are called **(pairwise) mutually exclusive** if  $A_i \cap A_j = \emptyset, \forall i, j$  with  $i \neq j$

Mutually

In general mutually exclusive events are such that no two events overlap

Exclusive

Suppose the function  $P$  associates a real value,  $P(A)$ , with each event  $A$  defined on a sample space  $\mathcal{S}$  such that

Definition  
Probability  
Set function,  
Axioms of  
Probability

1.  $0 \leq P(A)$  for every event  $A$

2.  $P(\mathcal{S}) = 1$

3. If  $A_1, A_2, \dots$  is a sequence of mutually exclusive events, then  $P(A_1 \cup \dots \cup A_n) = \sum_{i=1}^n P(A_i)$

then  $P$  is called a **probability set function** and  $P(A)$  is called the probability of  $A$

(1) - (3) are called the **Axioms of Probability**. These three axioms of probability are sufficient to allow a mathematical structure (the calculus of probability) to be developed.

**Note:** It does not matter if we assign the probabilities according to the classical, long-run relative frequency or subjective approach as long as the three axioms holds.

Suppose  $\mathcal{S}$  is a sample space for an experiment

Let  $A, B, C, A_1, A_2, \dots$  be events defined on  $\mathcal{S}$

Then:

Theorem  
Properties  
of Probabilities

1.  $0 \leq P(A) \leq 1$

$$P(A) = \sum_{a \in A} P(a) \leq \sum_{all\ a} P(a) = P(\mathcal{S}) = 1$$

2. **Probability of the Complement of an Event:**

$P(\bar{A}) = 1 - P(A)$  which implies  $P(\emptyset) = 1 - P(\mathcal{S}) = 0$

$$\sum_{all\ a} P(a) = \sum_{a \in A} P(a) + \sum_{a \in \bar{A}} P(a) \Rightarrow 1 = P(A) + P(\bar{A})$$

3. **Probability of a subset of an Event:**

If  $A \subset B$ , then  $P(A) \leq P(B)$

$$P(A) = \sum_{a \in A} P(a) \leq \sum_{a \in B} P(a) = P(B)$$

4. **Probability of the Union of Two Mutually Exclusive Events:**

Let  $A$  and  $B$  be mutually exclusive events

Then  $P(A \cup B) = P(A) + P(B)$

$$P(A \cup B) = \sum_{a \in A \cup B} P(a) = \sum_{a \in A} P(a) + \sum_{a \in B} P(a) = P(A) + P(B)$$

(Note: This works since for any  $a \in B, a \notin A$  and for any  $a \in A, a \notin B$ )

5. **Probability of the Union of Mutually Exclusive Events:**

Let  $A_1, \dots, A_n$  be mutually exclusive events

Then  $P(A_1 \cup \dots \cup A_n) = \sum_{i=1}^n P(A_i)$

$$A_i \cap A_j = \emptyset, i \neq j$$

6. For any two events  $A$  and  $B$ ,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A) + P(B) = P(A \cup B) + P(A \cap B)$$

(This is since when we add  $P(A)$  and  $P(B)$  we end up double counting the points at the intersection, once when we count  $P(A)$  and a second time when we count  $P(B)$ , so we subtract the intersection to make it count just once)

7. For any three events  $A, B$  and  $C$ ,

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

## Independent Events

Events  $A$  and  $B$  are **independent events** if and only if  $P(A \cap B) = P(A)P(B)$ . If they are not independent, we call the events **dependent**

Definition  
Two  
Independent  
Events

**Note:**  $P(A \cap B)$  can be written as  $P(AB)$

Independent events are such that the ratios

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

The events  $A_1, \dots, A_n$  are independent if and only if

$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k})$  for all sets  $(i_1, \dots, i_k)$  of distinct subscripts chosen from  $(1, \dots, n)$

Definition  
n  
Independent  
Events

It is not sufficient for events to just be pairwise independent. All groupings have to be independent.

$A$  and  $B$  are independent events if and only if  $\bar{A}$  and  $B$  are independent events. Similarly  $A$  and  $\bar{B}$  or  $\bar{A}$  and  $\bar{B}$

Theorem

**Proof:**

Note that  $B = (A \cap B) \cup (\bar{A} \cap B)$ , where  $A \cap B$  and  $\bar{A} \cap B$  are mutually exclusive events  
So  $P(B) = P(A \cap B) + P(\bar{A} \cap B)$ . Therefore,

$$P(\bar{A} \cap B) = P(B) - P(A \cap B) = P(B) - P(A)P(B) = P(B)(1 - P(A)) = P(\bar{A})P(B)$$



# Conditional Probability

The conditional probability of event  $A$ , given that event  $B$  occurs, is defined by

Definition

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

Conditional  
Probability

provided  $P(B) \neq 0$

Suppose  $A$  and  $B$  are two events such that  $P(A) > 0$  and  $P(B) > 0$ .  $A$  and  $B$  are independent if and only iff  $P(A|B) = P(A)$  or, equivalently, if  $P(B|A) = P(B)$

Theorem

For any two events  $A$  and  $B$ ,

Theorem

$$1. P(A \cap B) = P(A|B)P(B)$$

Multiplication

$$2. P(A \cap B) = P(B|A)P(A)$$

rule

Let the sample space,  $\mathcal{S}$ , be partitioned into  $k$  mutually exclusive sets  $B_1, B_2, \dots, B_k$  such that  $P(B_i) > 0$ , for  $i = 1, 2, \dots, k$  and  $\mathcal{S} = B_1 \cup \dots \cup B_k$ .

Theorem

Then, for any event  $A$ ,  $P(A) = \sum_{i=1}^k P(A \cap B_i)$

Law of

Total

Probability

Let the sample space,  $\mathcal{S}$ , be partitioned into  $k$  mutually exclusive sets  $B_1, \dots, B_k$  such that  $P(B_i) > 0$ , for  $i = 1, \dots, k$  and  $\mathcal{S} = B_1 \cup \dots \cup B_k$ .

Theorem

Then, for any events  $A$ , and for  $j = 1, \dots, k$ , we have

Bayes'

Theorem

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^k P(A|B_i)P(B_i)}$$

**Proof:**

$$P(B_j|A) = \frac{P(B_j \cap A)}{P(A)} = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^k P(A \cap B_i)} = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^k P(A|B_i)P(B_i)}$$

# Random Variables and Distributions

A random variable (e.g.  $X$ ) is a function that assigns a real number to each point in a sample space  $\mathcal{S}$

Definition

Random  
Variable

The event  $[X = 2]$  means  $\{s \in \mathcal{S} | X(s) = 2\}$  or the set of all points,  $s$ , in the sample space,  $\mathcal{S}$ , such that  $X(s) = 2$ .

We write the probability of this event as  $P(X = 2)$

The random variable,  $X$ , takes a finite or countably infinite number of distinct values (the range of  $X$  is countable)

Definition

Discrete  
Random

$X$  can take any value in a non-degenerate interval (range of  $X$  is not countable)

Variable  $X$

Definition

Continuous

The **function**  $f(x) = P(X = x)$ , for all  $x$  in the set of all possible values of  $X$

Random

Variable  $X$

Definition

1.  $f(x) \geq 0$  for all values of  $x$

Probability

2.  $\sum_{all\ x} f(x) = 1$  (sum of all probabilities is 1)

Function