STAT2402: Analysis of Observations

R. Nazim Khan
Department of Mathematics and Statistics
nazim.khan@uwa.edu.au

The University of Western Australia

Life lessons

The two most important days of your life are

the day you are born

and

the day your realise why.

Mark Twain

Week 3—Part I

- Revision of the linear statistical model
- 2. Introduction to Probability
- Random variables and discrete distributions

2. Introduction to Probability

- In this section we wil cover some basic probability.
- There is more here in the notes than we will cover in lectures. The rest is for your information and education.
- The concepts of conditional probability and independence are very important. We will cover the essential aspects. We may cover some examples but the rest is for you to look over.

Learning Outcomes

At the end of this chapter you should be able to:

1. use the axioms of probability to prove simple probability results;
2. use the rules of probability to solve probability problems;
3. understand the concept of conditional probability and use it to solve problems;
4. know the multiplication rule and use it to solve problems;
5. understand the concept of independence and disjoint (mutually exclusive) events, and use these to solve problems;
6. solve probability problems using tree diagrams.

Brief History of Probability

- A gambler's dispute in 1654 led to the creation of a mathematical theory of probability by two French mathematicians, Blaise Pascal and Pierre de Fermat.
- Antoine Gombaud (who called himself Chevalier de Méré, meaning Knight of Mé'ré), a French noble man and a writer with an interest in gaming and gambling questions, called Pascal's attention to an apparent contradiction concerning a popular dice game.

The problem

- The game consisted of throwing a pair of dice 24 times. The problem was to decide whether or not to bet even money on the occurrence of at least one double six during the 24 throws. A seemingly well-established gambling rule led de Méré to believe that betting on a double six in 24 throws would be profitable. but his own calculations showed just the opposite.
- This problem posed by de Méré and others led to an exchange of letters between Pascal and Fermat in which the fundamental principles of probability were formulated for the first time.
- **Question**: What is the probability of winning in the above game if you bet on "at least one double six"?

Contents

2.1 Notation and Terminology

What is random?

- When the outcome cannot be predicted with certainty in advance.

A **random experiment** is a process that generates outcomes that can only be described in **terms of probabilities**.

- Tossing a coin
- Rolling a die

Sample Space

A *sample space*, denoted S, is the set of all possible outcomes of a random experiment.

Example 2.1 If a coin is tossed twice then the sample space is

$$\{HH, HT, TH, TT\}$$

Any outcome in the sample space is called an *elementary event*. For example, $\{HH\}$ is an elementary event.

An *event* is a subset of the sample space. For example, $\{HH, HT, TH\}$ is an event, which can be described is words as *at least one H*.

Assigning probabilities

The probability of an event A is the number of outcomes in the sample space favourable to A divided by the total number of outcomes, that is

$$P(A) = \frac{\text{\# of outcomes favourable to } A}{\text{Total number of outcomes}}$$

Example 2.1 (ctd) Tossing a coin twice.

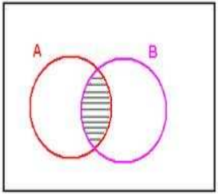
$$\begin{aligned} P(\text{2H}) &= \frac{1}{4} \\ P(\text{At least one H}) &= \frac{3}{4} \\ P(\text{No H}) &= \frac{1}{4}. \end{aligned}$$

The null or empty event

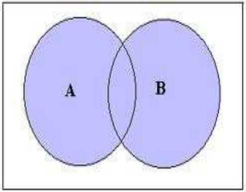
The *null* or *empty event* is denoted ϕ (phi) and contains no outcomes.

Example 2.1 (ctd) Toss a coin twice. The event “Obtaining 3 H” is a null event — we cannot get three H from two tosses of a coin.

Union and Intersection

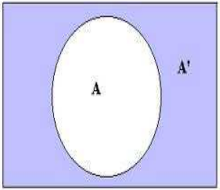


Intersection $A \cap B$ is the set of points that are both in A and in B.

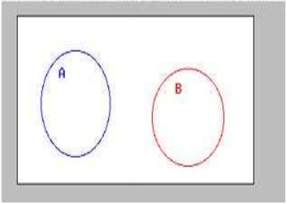


Union $A \cup B$ is the set of points that are in A or B or both.

Complementary and Disjoint Events



The *complement* of an event A, denoted A', $\overline{A}$ or A^c, is the *set of points which are not in* A.



A and B are *mutually exclusive* or *disjoint* if $A \cap B = \phi$. That is, A and B have no intersection.

Exercises

1. What is the complement of A^c?
2. What is the intersection of A and A^c?

Example

Two coins are tossed. Let A denote the event that the first toss yields a H.

(a) What is the complement of A?

$$S = \{HH, HT, TH, TT\}, \quad A = \{HH, HT\}, \quad A^c = \{TH, TT\}$$

(b) Write down an event that is disjoint with A.
A^c is disjoint with A. $\{TH\}$ and $\{TT\}$ are also disjoint with A.

2.2 Axioms of Probability

Kolmogorov (1903-1987), one of the most influential mathematicians of the twentieth century, put probability on a firm mathematical foundation (1933). He postulated the three axioms of probability.

P1. $P(S) = 1$ (Something must happen)

P2. $P(A) \geq 0$ for any event A

P3. If A and B are disjoint then

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

Exercise Draw a Venn diagram to represent Axiom P3.

Reference:

http://www.kolmogorov.com/

2.3 Rules of Probability

Based on the Axioms of Probability, we can derive the following rules.

P1. $P(\phi) = 0$

P2. $P(A^c) = 1 - P(A)$

P3. $0 \leq P(A) \leq 1$

P4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Proof

P1 $P(\phi) = 0$. Note that S and ϕ are disjoint, since $S \cap \phi = \phi$. also, $S \cup \phi = S$ and $P(S) = 1$, so

$$\begin{aligned} P(S \cup \phi) &= P(S) \\ \text{so } P(S) + P(\phi) &= P(S) \\ 1 + P(\phi) &= 1 \\ \Rightarrow P(\phi) &= 0. \end{aligned}$$

P2 $P(A^c) = 1 - P(A)$. A^c and A are disjoint since $A \cap A^c = \phi$. Also, $A \cup A^c = S$. Then

$$\begin{aligned} P(A \cup A^c) &= P(S) \\ \text{so by P3 and P1 } P(A) + P(A^c) &= 1 \\ \Rightarrow P(A^c) &= 1 - P(A). \end{aligned}$$

Note that it also follows that $P(A) = 1 - P(A^c)$.

Proof (ctd)

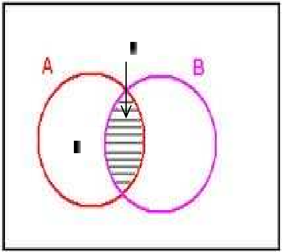
P3 $0 \leq P(A) \leq 1$. By P2, $P(A) \geq 0$, or $0 \leq P(A)$. Also by P2, $P(A^c) \geq 0$. But

$$P(A) = 1 - \underbrace{P(A^c)}_{\geq 0} \leq 1.$$

It follows that $0 \leq P(A) \leq 1$.

P4

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
Exercise. Note that
$P(A) = P(A \cap B) + P(A \cap B^c)$.
This is sometimes read as $P(A)$ equals probability of A with B plus probability of A without B.
This is a form of the *Theorem of Total Probabilities*. Note that



Example 2.3

Let $P(A) = 0.4$, $P(B) = 0.5$ and $P(A \cap B) = 0.3$. Calculate

(a) $P(\text{at least one of } A \text{ and } B \text{ occurs})$,

(b) $P(\text{only } A \text{ occurs})$,

(c) $P(\text{neither } A \text{ nor } B \text{ occurs})$

Solution

(a) $P(A \cup B) = P(A) + P(B) - PA \cap B) = 0.4 + 0.5 - 0.3 = 0.6$.

(b) $P(A) = P(A \cap B) + P(A \cap B^c)$, so
$P(A \cap B^c) = P(A) - P(A \cap B) = 0.4 - 0.3 = 0.1$.

(c) Neither A nor B occurs is equivalent to
$(A \cup B)^c = 1 - P(A \cup B) = 1 - 0.6 = 0.4$. Draw a Venn diagram to illustrate this situation.

2.4 Conditional Probability

For events A and B, the probability of A occurring when it is known that B *has occurred* is called the *conditional probability of A given B*, denoted

$$P(A \mid B),$$

read as "probability of A given B." Note that $P(A \mid B)$ is a probability, so it satisfies *ALL* the rules of a probability.

Definition If A and B are two events such that $P(B) > 0$, then

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

Since B has occurred, now the probability of A occurring is that proportion of B where A occurs.

Interpretation

$P(A \mid B)$ is the probability of A under the reduced sample space corresponding to B having occurred.

Example 2.4: Bank Data

Consider the example below, listing the workers by gender and job-grade from the Bank data.

	Job Grade						
	1	2	3	4	5	6	Total
Female	48	29	36	17	9	1	140
Male	12	13	7	11	12	13	68
Total	60	42	43	28	21	14	208

Example 2.4 (ctd)

Calculate the following probabilities.

(a) $P(\text{Male} \mid \text{Job Grade} \geq 5) = \frac{25}{35} = \frac{5}{7} \approx 0.7143$. There are proportionally more males at the higher job grades.

(b) $P(\text{Job Grade} \geq 5 \mid \text{Female}) = \frac{10}{140} = \frac{1}{14} \approx 0.0714$

(c) $P(\text{Female} | \text{Job Grade} \leq 4) = \frac{130}{173} \approx 0.7514$. There are proportionally more females at the lower job grades.

2.5 Multiplication Rule

The conditional probability rule can be written as

Multiplication Rule

$$P(A \cap B) = P(A)P(B \mid A) = P(B)P(A \mid B),$$

provided $P(A) > 0$ and $P(B) > 0$.

Notes

If A and B happen together, then

- first A happens and then B happens given that A has happened, **or**
- first B happens and then A happens given that B has happened.

2.6 Tree Diagrams

Useful in solving conditional probability problems.

Example 2.5

Machine A and B turn out respectively 10% and 90% of the total production of a certain type of article. The probability that Machine A produces a defective article is 0.01, while that for machine B is 0.05. What is the probability that an article taken at random from a days production was made by Machine A, given that the article is faulty?

Solution 1. Define appropriate events.
Let A be the event that that article is produced by machine A and D be the event that the article is defective.
2. Express the given probabilities in terms of these events and draw a tree diagram if required.

Example 2.5 (ctd)

Example 2.5 (ctd)

3. Express the required probability in terms of the defined events.

$$
\begin{aligned}
P(A \mid D) &= \frac{P(A \cap D)}{P(D)} \\
&= \frac{P(A \cap D)}{P(A \cap D) + P(\overline{A} \cap D)} \\
&= \frac{0.1 \times 0.01}{0.1 \times 0.01 + 0.9 \times 0.05} \\
&= \frac{1}{46} \approx 0.0217.
\end{aligned}
$$

Example 2.6: Bank Data

Consider again the data of Example 3.4.

	Job Grade						
	1	2	3	4	5	6	Total
Female	48	29	36	17	9	1	140
Male	12	13	7	11	12	13	68
Total	60	42	43	28	21	14	208

What is the probability that

(a) an employee is female and is at job grade 5 or above?

$= 10/208 \approx 0.0481$

(b) an employee at job grade 4 or below is male?

$= 43/173 \approx = 0.2486.$

Example 2.6 (ctd)

(c) an employee at job grade 5 or above is female?

$= 10/35 \approx 0.2857$

(d) an employee at job grade 5 or above is male?

$= 1 - 0.2857 = 0.7143$

On the basis of the above calculations, what can be concluded about the promotion culture in the bank?
The chance of being at a higher job grade for females is less than 30%, indicating that there is a gender bias in promotions. BUT, what other variables affect promotion: education, experience, ...

2.7 Independent Events

Intuitively, two events A and B are independent if the occurrence of one does not affect the probability of the occurrence of the other.

Formally, events A and B are independent if

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

Otherwise they are dependent.
Note To determine if two events are independent, the above condition in equation (1) needs to be verified.

Notes

1. If A and B are independent then

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

   so the occurrence of B does not affect the probability of the occurrence of A.

2. Similarly,

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

   so the occurrence of A does not affect the probability of the occurrence of B.

This is intuitively what independence means. That is, if two events are independent then the occurrence of one does not affect the probability of the occurrence of the other.

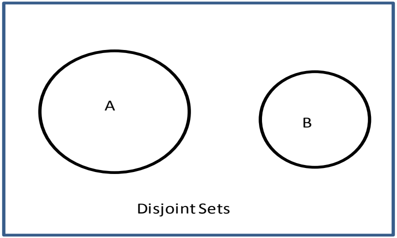
Do not confuse mutually exclusive and independence!!

A and B are mutually exclusive (disjoint) if

$$P(A \cap B) = 0.$$

A and B are independent if

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



Example 2.7: Bank Data

Consider again the data of Example 3.4.

	Job Grade						
	1	2	3	4	5	6	Total
Female	48	29	36	17	9	1	140
Male	12	13	7	11	12	13	68
Total	60	42	43	28	21	14	208

(a) Let A be the event that an employee is male, and B be the event that an employee is at job grade 5 or above. Are the events A and B independent?

Example 2.7 (ctd)

From the table,

$$P(A) = 68/208, P(B) = 35/208, P(A \cap B) = 25/208 \approx 0.1208,$$

so

$$P(A) \times P(B) = \frac{68}{208} \times \frac{35}{208} \approx 0.0551 \neq P(A \cap B).$$

Thus A and B are NOT independent.

(b) Is gender independent of job grade? If not, how what is the relationship between gender and job grade?

Gender and job grade are dependent by part (a) above. Males are more likely to be at higher job grades and females are more likely to be at lower job grades.

Example 2.8

A woman is selling her house. She believes that there is 0.3 chance that a person who inspects her house will purchase it. Assuming that the people inspecting the house decide independently whether or not to purchase the house, what is the probability that more than two people will have to inspect the house before it is sold?

Solution

Let A_1 denote the event that the first person to view the house buys it, and A_2 denote the event that the second person to view the house buys it. The A_1 and A_2 are independent. We need $P(\overline{A}_1 \cap \overline{A}_2)$, that is, the first two people to view the house do not buy it.

Example 2.8 (ctd)

Result If A and B are independent events then the events
A and $\overline{B}$, $\overline{A}$ and B, and $\overline{A}$ and $\overline{B}$
are also independent.

Proof Exercise.
By the above result,

$$P(\overline{A}_1 \cap \overline{A}_2) = P(\overline{A}_1) \times P(\overline{A}_2) = (1 - 0.3) \times (1 - 0.3) = 0.49.$$

Exercise Draw a tree diagram for this problem.

Example 2.9

While searching for oil in Australia, an oil explorer orders seismic tests to determine if oil is likely to be found in a certain drilling area. The following probabilities summarise past results concerning the reliability of the test: when oil does exist in the testing area, the test will indicate so 85% of the time; when oil does not exist in the testing area, the probability is 0.03 that the test will erroneously indicate that oil does exist. Preliminary exploration by geologists indicates that the probability of the existence of oil deposits in the test area is 0.45. If a seismic test is conducted and indicates the presence of oil, what is the probability that an oil deposit really exists?

Solution

Let O denote the event that oil is present, and T denote the event that the test indicates the presence of oil.

Example 2.9 (ctd)

2.8 Summary

1. Know the axioms of probability.
2. Know the rules of probability.
3. Multiplication rule.
4. Independent and mutually exclusive (disjoint) events.
5. Solve problems in conditional probability using tree diagrams.