CHEM. ENG. II : PROBABILITY AND STATISTICS

Chapter 1. Basic Probability Concepts

The random variation associated with "measurement" procedures in a scientific analysis requires a framework in which the **uncertainty** and **variability** that are inherent in the procedure can be handled.

1.1 Experiments and Events

An **experiment** is any procedure

(a) with a well-defined **set** of possible outcomes - the **sample space**, S.
(b) whose **actual** outcome is not known in advance.

A **sample outcome**, s, is precisely one of the possible outcomes of the experiment.

The **sample space**, S, is the entire set of possible outcomes.

EXAMPLES:

(a) Coin tossing: $S = \{H, T\}$.
(b) Dice : $S = \{1, 2, 3, 4, 5, 6\}$.
(c) Proportion material content: $S = \{x : 0 \leq x \leq 1\}$
(d) Failure time measurement: $S = \{x : x > 0\} = \mathbb{R}^+$
(e) Temperature measurement: $S = \{x : a \leq x \leq b\} \subseteq \mathbb{R}$

There are two basic types of experiment, namely

COUNTING

and

MEASUREMENT

- we shall see that these two types lead to two distinct ways of specifying probability distributions.

The collection of sample outcomes is a **set**, so we write

$$s \in S$$

if s *is a member of the set* S.

DEFINITION

An **event** E is a set of the possible outcomes of the experiment, that is E is a **subset** of S, $E \subseteq S$, E **occurs** if the actual outcome is in this set.

NOTE: the sets S and E can be either **countable**, that is, can be written as a list of items, for example,

$$E = \{s_1, s_2, ..., s_n, ...\}$$

which may a finite or infinite list, or **uncountable**, that is, can only be represented by a continuum of outcomes, for example

$$E = \{x : 0.6 < x \leq 2.3\}$$

Events are manipulated using **set theory** notation; if E, F are two events, $E, F \subseteq S$,

Union	$E \cup F$	"E **or** F **or both** occurs"
Intersection	$E \cap F$	"E **and** F occur"
Complement	E'	"E **does not** occur"

We can interpret the events $E \cup F$, $E \cap F$, and E' in terms of collections of sample outcomes, and use Venn Diagrams to represent these concepts.

Special cases of events:

THE IMPOSSIBLE EVENT	$\emptyset$	the emptyset, the collection of sample outcomes with zero elements
THE CERTAIN EVENT	Ω	the collection of all sample outcomes

DEFINITION

Events E and F are **mutually exclusive** if

$$E \cap F = \emptyset$$

that is, the collections of sample outcomes E and F have no element in common.

1.2 Results in events manipulation

For events E, F, and G, the following equations can be used to simplify complex expressions;

ASSOCIATIVITY	$(E \cup F) \cup G = E \cup (F \cup G)$ $(E \cap F) \cap G = E \cap (F \cap G)$
DISTRIBUTIVITY	$E \cup (F \cap G) = (E \cup F) \cap (E \cup G)$ $E \cap (F \cup G) = (E \cap F) \cup (E \cap G)$
also	$(E \cup F)' = E' \cap F'$, $(E \cap F)' = E' \cup F'$

DEFINITION

Events $E_1, ..., E_k$ form a **partition** of event $F \subseteq S$ if

(a) $E_i \cap E_j = \emptyset$ for all i and j (b) $\bigcup_{i=1}^{k} E_i = E_1 \cup E_2 \cup ... \cup E_k = F$.

We are interested in mutually exclusive events and partitions because when we carry out probability calculations we will essentially be counting or enumerating sample outcomes; to ease this counting operation, it is desirable to deal with collections of outcomes that are completely distinct or **disjoint**.

1.3 The rules of probability

The probability function P(.) is a set function that assigns weight to collections of sample outcomes. We can consider assigning probability to an event by adopting

CLASSICAL APPROACH	consider equally likely outcomes
FREQUENTIST APPROACH	consider long-run relative frequencies
SUBJECTIVE APPROACH	consider your personal degree of belief

It is legitimate to use any justification where appropriate or plausible.

Mathematical Properties - The Probability Axioms

It is sufficient to require that the set function P(.) must satisfy the following properties.

For any events E and F in sample space S,

(1) $0 \leq \mathrm{P}(E) \leq 1$
(2) $\mathrm{P}(\Omega) = 1$
(3) If $E \cap F = \emptyset$, then $\mathrm{P}(E \cup F) = \mathrm{P}(E) + \mathrm{P}(F)$

Corollaries :
$\mathrm{P}(E') = 1 - \mathrm{P}(E)$, $\mathrm{P}(\emptyset) = 0$
If $E_1, ..., E_k$ are events such that $E_i \cap E_j = \emptyset$ for all i, j, then

$$\mathrm{P}\left(\bigcup_{i=1}^{k} E_i\right) = \mathrm{P}(E_1) + \mathrm{P}(E_2) + \ldots + \mathrm{P}(E_k).$$

If $E \cap F \neq \emptyset$, then $\mathrm{P}(E \cup F) = \mathrm{P}(E) + \mathrm{P}(F) - \mathrm{P}(E \cap F)$

EXAMPLE CALCULATION Examination Pass Rates

The examination performance of students in a year of eight hundred students is to be studied: a student either chooses an essay paper or a multiple choice test. The pass figures and rates are given in the table below:

	PASS	FAIL	PASS RATE
FEMALE	200	200	0.5
MALE	240	160	0.6

The result of this study is clear: the pass rate for MALES is higher than that for FEMALES.

Further investigation revealed a more complex result: for the essay paper, the results were as follows;

	PASS	FAIL	PASS RATE
FEMALE	120	180	0.4
MALE	30	70	0.3

so for the essay paper, the pass rate for FEMALES is higher than that for MALES.

For the multiple choice test, the results were as follows;

	PASS	FAIL	PASS RATE
FEMALE	80	20	0.8
MALE	210	90	0.7

so for the multiple choice paper, the pass rate for FEMALES is higher than that for MALES.

Hence we conclude that FEMALES have a higher pass rate on the essay paper, and FEMALES have a higher pass rate on the multiple choice test, but MALES have a higher pass rate overall.

This apparent contradiction can be resolved by careful use of the probability definitions. First introduce notation; let E be the event that the student chooses an essay, F be the event that the student is female, and G be the event that the student passes the selected paper.

Exercise: Draw a Venn diagram to represent this problem.

1.4 Conditional probability

DEFINITION

For two events E and F with $\mathrm{P}(F) > 0$, the **conditional probability** that E occurs, **given** that F occurs, is written $\mathrm{P}(E|F)$, and is defined by

$$\mathrm{P}(E|F) = \frac{\mathrm{P}(E \cap F)}{\mathrm{P}(F)}$$

so that

$$\mathrm{P}(E \cap F) = \mathrm{P}(E|F)\mathrm{P}(F)$$

It is easy to show that this new probability operator $\mathrm{P}(\ .\ |\ .\ )$ satisfies the probability axioms.

[In the exam results problem, what we really have specified are conditional probabilities. From the pooled table, we have

$$\mathrm{P}(G|F) = 0.5 \qquad \mathrm{P}(G|F') = 0.6,$$

from the essay results table, we have

$$\mathrm{P}(G|E \cap F) = 0.4 \qquad \mathrm{P}(G|E \cap F') = 0.3,$$

and from the multiple choice table, we have

$$\mathrm{P}(G|E' \cap F) = 0.8 \qquad \mathrm{P}(G|E' \cap F') = 0.7$$

and so interpretation is more complicated than originally thought.]

The probability of the **intersection** of events $E_1, ..., E_k$ is given by the **chain rule**

$$\mathrm{P}(E_1 \cap ... \cap E_k) \ = \ \mathrm{P}(E_1)\mathrm{P}(E_2|E_1)\mathrm{P}(E_3|E_1 \cap E_2)...\mathrm{P}(E_k|E_1 \cap E_2 \cap ... \cap E_{k-1})$$

Special Case: Independence

Events F and F are **independent** if

$$\mathrm{P}(E|F) \ = \ \mathrm{P}(E) \text{ so that } \mathrm{P}(E \cap F) \ = \ \mathrm{P}(E)\mathrm{P}(F)$$

and so if $E_1, ..., E_k$ are independent events, then

$$\mathrm{P}(E_1 \cap ... \cap E_k) = \prod_{i=1}^{k} \mathrm{P}(E_i) = \mathrm{P}(E_1)...\mathrm{P}(E_k)$$

1.5 The Theorem of Total Probability

THEOREM

If events $E_1, ..., E_k$ form a partition of event $F \subseteq S$, and event $G \subseteq S$ is such that $\mathrm{P}(G) > 0$, then

$$\mathrm{P}(F) \ = \sum_{i=1}^{k} \mathrm{P}(F|E_i)\mathrm{P}(E_i)$$

$$\mathrm{P}(F|G) \ = \sum_{i=1}^{k} \mathrm{P}(F|E_i \cap G)\mathrm{P}(E_i|G)$$

Proof

We have by assumption that

$$F = \bigcup_{i=1}^{k} (E_i \cap F) \Longrightarrow \mathrm{P}(F) = \sum_{i=1}^{k} \mathrm{P}(E_i \cap F) = \sum_{i=1}^{k} \mathrm{P}(F|E_i)\mathrm{P}(E_i)$$

by probability axiom (3), as the collection $E_1 \cap F, ..., E_k \cap F$ are mutually exclusive.

Exercise: Attempt to resolve the examinations results paradox using this Theorem.

1.6 Bayes Theorem

THEOREM

For events E and F such that $\mathrm{P}(E)$, $\mathrm{P}(F) > 0$,

$$\mathrm{P}(E|F) = \frac{\mathrm{P}(F|E)\mathrm{P}(E)}{\mathrm{P}(F)}$$

If events $E_1, ..., E_k$ form a partition of S, with $\mathrm{P}(E_i) > 0$ for all i, then then

$$\mathrm{P}(E_i|F) = \frac{\mathrm{P}(F|E_i)\mathrm{P}(E_i)}{\mathrm{P}(F)} = \frac{\mathrm{P}(F|E_i)\mathrm{P}(E_i)}{\sum_{j=1}^{k}\mathrm{P}(F|E_j)\mathrm{P}(E_j)}$$

Proof

We have from the conditional probability definition that

$$\mathrm{P}(E \cap F) = \mathrm{P}(E|F)\mathrm{P}(F) \qquad \text{and} \qquad \mathrm{P}(E \cap F) = \mathrm{P}(F|E)\mathrm{P}(E)$$

and hence equating the right hand sides of the two equations we have

$$\mathrm{P}(E|F)\mathrm{P}(F) = \mathrm{P}(F|E)\mathrm{P}(E)$$

and hence the result follows.

Note that in the second part of the theorem,

$$\mathrm{P}(E_i|F) = \frac{\mathrm{P}(F|E_i)\mathrm{P}(E_i)}{\mathrm{P}(F)} = \frac{\mathrm{P}(F|E_i)}{\mathrm{P}(F)}\,\mathrm{P}(E_i)$$

so the probabilities $\mathrm{P}(E_i)$ are re-scaled to $\mathrm{P}(E_i|F)$ by conditioning on F. Note that

$$\sum_{i=1}^{k} \mathrm{P}(E_i|F) = 1$$

This theorem is very important because, in general,

$$\mathrm{P}(E|F) \neq \mathrm{P}(F|E)$$

and it is crucial to condition on the correct event in a conditional probability calculation.

EXAMPLE Lie-detector test.

In an attempt to achieve a criminal conviction, a lie-detector test is used to determine the guilt of a suspect. Let G be the event that the suspect is guilty, and let T be the event that the suspect fails the test.

The test is regarded as a good way of determining guilt, because laboratory testing indicate that the detection rates are high; for example it is known that

$$\mathrm{P}[\text{ Suspect Fails Test } | \text{ Suspect is Guilty }] = \mathrm{P}(T|G) = 0.95 = 1 - \alpha, \text{ say}$$

$$\mathrm{P}[\text{ Suspect Passes Test } | \text{ Suspect is Not Guilty }] = \mathrm{P}(T^{'}|G^{'}) = 0.99 = \beta, \text{ say.}$$

Suppose that the suspect fails the test. What can be concluded ? The probability of real interest is $\mathrm{P}(G|T)$; we do not have this probability but can compute it using Bayes Theorem.

For example, we have

$$P(G|T) = \frac{P(T|G)P(G)}{P(T)}$$

where $P(G)$ is not yet specified, but $P(T)$ can be computed using the Theorem of Total probability, that is,

$$P(T) = P(T|G)P(G) + P(T|G^{'})P(G^{'})$$

so that

$$P(G|T) = \frac{P(T|G)P(G)}{P(T|G)P(G) + P(T|G')P(G')}$$

Clearly, the probability $P(G)$, the probability that the suspect is guilty *before* the test is carried out, plays a crucial role. Suppose, that $P(G) = p = 0.005$, so that only 1 in 200 suspects taking the test are guilty. Then

$$P(T) = 0.95 \times 0.005 + 0.01 \times 0.995 = 0.0147$$

so that

$$P(G|T) = \frac{0.95 \times 0.005}{0.95 \times 0.005 + 0.01 \times 0.995} = 0.323$$

which is still relatively small. So, as a result of the lie-detector test being failed, the probability of guilt of the suspect has increased from 0.005 to 0.323.

More extreme examples can be found by altering the values of α, β and p.

Exercise: Find the general relationship between α, β, p and $p^* = P(G|T)$.