1
Introduction

1.1 The nature of statistical prediction analysis

An essential feature of statistical prediction analysis is that it involves two experiments e and f. From the information which we gain from a performance of e, the *informative* experiment, we wish to make some reasoned statement concerning the performance of f, the *future* experiment. In order that e should provide information on f there must be some link between these two experiments. Throughout this book we shall deal with problems where this link is through the indexing parameter of the two experiments e and f, and so we make the following assumption.

> *Assumption 1* The class of probability models which form the possible descriptions of e and the class of possible models for f have the same index set Θ, and the true models have the same (though unknown) index θ^*.

A further general feature of all the problems we shall consider is contained in the following independence assumption.

> *Assumption 2* For given index θ the experiments e and f are independent.

By adopting this second assumption we deliberately exclude a range of prediction problems in which f is a continuation of some stochastic process of which e records a realisation to date. Techniques such as forecasting by exponential weighting, linear least squares prediction and time series analysis are thus outside the scope of this book.

To give some idea of the wide applicability of statistical prediction analysis as defined above and to motivate the development of appropriate theory we devote the remainder of this chapter to the presentation of specific prediction problems. All these problems are later analysed and extended in the sections indicated in the text.

1.2 Some examples

In its most direct form statistical prediction analysis may simply be the

provision of some probabilistic statement about the likely outcome of the performance of f.

Table 1.1 *Survival times (weeks) of 20 carcinoma patients*

25	45	238	194	16	23	30	16	22	123
51	412	45	162	14	72	5	35	30	91

Example 1.1

Medical Prognosis. The data of table 1.1 are the survival times (weeks) of 20 patients presenting with a certain type of carcinoma and receiving treatment of preoperative radiotherapy followed by radical surgery. On the basis of this information what can appropriately be said about the future of a new patient with this type of carcinoma and assigned to this form of treatment? Clearly any rational statement would regard 100 weeks survival as much more plausible than 500 weeks survival, but how should such views be summarised and quantified? What is a reasonable assessment of the probability that the patient will survive 100 weeks?

In this example the informative experiment e consists of recording the survival times of the 20 patients already treated. The future experiment f consists of treating the new patient similarly and recording his survival time. If no change in the treatment has been made since the conducting of e, then e and f consist respectively of 20 replicates and a single replicate of the same basic trial (record the survival time of a treated patient) and are independent. Assumptions 1 and 2 are therefore satisfied.

Attempts to quantify medical prognoses are of vital importance when similar information on an alternative treatment, for example radical surgery followed by postoperative radiotherapy, is available and a choice has to be made between treatments for a particular patient.

A detailed analysis and developments of this example are given in §2.6.

There are many less direct forms of statistical analysis than that of providing probabilistic statements. For example the problem may be one of choosing between alternative courses of action. If the consequences of taking a course of action depend on the outcome of f, then we still technically describe the problem as one of statistical prediction analysis.

Example 1.2

Machine tool replacement. Table 1.2 shows the recorded lifetimes of 24 machine tools of a certain type. In a factory using one of these machine tools the question of the best inspection and replacement policy is under discussion. If the tool wears out while unattended there is a loss of $\xi = 1.8$ per minute until such time as it is inspected (and immediately replaced). To have an inspector in attendance at the machine tool costs $\eta = 2.4$ per minute. If the

Table 1.2 *Lifetimes (minutes) of 24 machine tools*

47	62	111	47	57	14	290	118	19	4	17	46
5	239	9	140	89	94	217	35	103	28	37	111

tool is replaced before it wears out an overhead cost $\zeta = 54$ is incurred and also a loss at rate ξ per minute of its unused lifetime has to be debited. What is the optimum policy on the basis of all the information?

Here again the experiments e and f are easily identified. If we regard a basic trial as consisting of the recording of the lifetime of a single machine tool then e and f consist of 24 replicates and a single replicate of the basic trial, and satisfy assumptions 1 and 2. Again we clearly wish to infer something about the performance of f (the current machine tool) from the information contained in e. But a statement about the relative plausibilities of the possible outcomes of f is not sufficient. We must decide on one of many possible courses of action, namely the time periods during which we wish an inspector to be present to investigate whether the tool is still functioning or has already broken down. Suppose for simplicity that the only courses of action open to us are to select a time, a say, at which to send in an inspector with instructions to replace the tool immediately. The consequences of taking action a depend on the outcome of f, the actual lifetime y of the current machine tool. For if a exceeds y there is lost production time $a - y$ with a corresponding loss of $\xi(a - y)$, whereas if y exceeds a, the overhead scrapping loss ζ is incurred together with a debit of $\xi(y - a)$ for unused productive capacity. Thus a prediction associated with the performance of the future experiment f is necessary for any rational analysis of the problem but the prediction is a means to the end of selecting an appropriate course of action. We have thus here a less direct form of statistical prediction analysis than in our previous example. The analysis of this problem is developed in §§3.2, 3.4.

Example 1.3

A quality control problem. Items are produced independently in large batches by a firm. The items may be either effective or defective and it is recognised by both manufacturer and customer that batches vary considerably in the number of effectives they contain. The terms of a suggested contract between manufacturer and customer require the manufacturer to test destructively 5 of the components of each batch. The remainder of the batch is to be supplied in packets of 25 with an accompanying statement about the maximum number of defectives each packet contains. The contract further requires that for at least 90 per cent of such batches the statement will be true for at least 80 per cent of packets. What statement strategy will fulfil the terms of this contract?

Here the future experiment f envisaged is the observation of the number of defectives in a packet of 25 components from a batch. The information that

we have available consists of the observation of the number x of defectives in a sample of 5 components.

How we use the information in e to meet the requirements on f and how we interpret the 90 per cent and 80 per cent in the statements are discussed later in §6.3.

In the examples so far discussed we have attempted to illustrate some of the basic structure of statistical prediction analysis. The main feature emerging is the need to make some prediction of the outcome y of a future experiment f based on the outcome x of an informative experiment e. The relevance of the information obtained from e to the future experiment f is contained in what can conveniently be called the predictive density function. This concept is dealt with in detail in chapter 2 and is central to much of the subsequent analysis. For example the predictive density function provides the quantification of medical prognosis that we seek for example 1.1. In chapter 3 the introduction of utility functions to quantify the measures of gains or losses involved in making a prediction enables decisive prediction problems such as example 1.2 to be analysed in detail. Informative prediction problems in which no such measures are available, as in example 1.3, are dealt with in chapters 4, 5 and 6. A substantial part of these chapters presents the theory of tolerance regions from a fresh viewpoint. Some interesting relationships between decisive and informative prediction are developed in chapter 7, which also reviews other approaches such as empirical Bayes and distribution-free prediction. The remainder of the book is then devoted to specific areas of application and particularly to even more indirect forms of statistical prediction analysis. Some of these forms are now illustrated by examples.

1.3 Examples of choice of future experiment

There are many problems in which there is a whole class F of possible future experiments and the problem is to determine which future experiment f satisfies certain desirable properties. In choosing this experiment we have to envisage its performance and for this reason such problems fall within the scope of statistical prediction analysis. We consider here two representative examples.

Example 1.4

A problem of laminate design. In the manufacture of a laminate several sheets of material are superimposed. The sheets are liable to contain flaws and the total number y of flaws in the finished product can be measured by an X-ray device. The durability of the product increases as the number t of component sheets increases, but there is an upper limit y_0 to the total number of flaws which can be allowed before the product is rejected.

Table 1.3 *Numbers of flaws in nine laminate specimens*

Number of sheets	5	6	7	8	9	10	11	12	13
Number of flaws	3	2	4	5	7	6	7	7	8

In a pilot experiment 9 specimens of the product were made with 5, 6, 7, ... , 13 sheets of material superimposed with resulting flaws as shown in table 1.3. The management wishes to market as durable a product as possible but has decided that at most $y_0 = 7$ flaws can be allowed in any product. If the profit per component sheet for accepted products is ten times the cost per component sheet for rejected products how many sheets should be superimposed?

In order to resolve this problem we have to consider the number of flaws which may result if we use t sheets. Thus we are forced to envisage a whole class of possible experiments

$$F = \{f_t : t = 1, 2, \dots\},$$

where f_t denotes the experiment of counting the total number of flaws in a laminate of t sheets. The problem is then to choose which future experiment f_t gives as high durability as possible and yet attempts to meet the flaw limitation. The information available comes from an informative experiment e yielding the data of table 1.3 and which could formally be written in the form

$$e = \{f_5, f_6, \dots, f_{13}\},$$

a set of independent performances of $f_5, f_6, \dots, f_{13}$. The direct prediction problem for f_t enters into our attempts to balance our desire to increase t (and hence the durability) and our concern that the number of flaws may increase beyond the acceptable limit. This *regulation* example is analysed in §9.3.

Example 1.5

Maximising yield of an industrial process. The yields (kg) shown in table 1.4 were obtained in an experiment in which an industrial process was run successively at 5 different temperatures and 3 different pressures, each combination of temperature and pressure being used twice. What combination of temperature and pressure should be used in order to maximise the yield in a future run of the process?

Here the problem is to determine at what combination $t = (t_1, t_2)$ of temperature t_1 and pressure t_2 to run the process. Denote by f_t or f_{t_1, t_2} the future experiment which records the yield from an operation of the industrial process at temperature t_1 and pressure t_2. We are thus forced to consider the class of future experiments

Table 1.4 *Yields (kg) from 30 process runs at different temperature-pressure combinations*

Temperature (°C)	Pressure (atmospheres)		
	1.00	1.25	1.50
50	65, 68	70, 72	73, 74
60	72, 70	75, 75	77, 76
70	73, 75	81, 83	79, 78
80	76, 75	81, 79	75, 77
90	76, 76	78, 80	76, 73

$$F = \{f_t : t \in T\},$$

where T denotes the set of possible temperature–pressure combinations. The informative experiment consists of 30 independent experiments (process runs) each of f_t type. To choose an experiment from the set F we must again envisage the outcomes of such future experiments and so are involved in statistical prediction analysis. This *optimisation* problem is analysed in §9.5.

1.4 Examples of detection of future experiment

A common statistical problem is to detect which one of a class F of 'future' experiments has already been performed from the information from e and the known outcome y of the performed future experiment. While in such circumstances it may seem strange to use the term statistical prediction analysis we shall see that we are led inevitably to the same concepts of prediction as we have already encountered. Indeed we have to envisage prediction for each of the possible future experiments in F. Two examples illustrate the nature of the problem here.

Example 1.6

Antibiotic assay. When a droplet of specified volume of an antibiotic is placed on an infected medium on a Petri dish and kept under controlled conditions the antibiotic clears a circular area of the medium. Moreover the diameter of the cleared area depends on the concentration of the antibiotic although this relationship is not a deterministic one. The idea underlying a biological assay of the *unknown* concentration of antibiotic in a blood specimen from a patient is to place droplets of standard antibiotic at different *known* concentrations and droplets from the specimen on the same batch of infected medium (fig. 1.1). The problem is then to infer from the relative sizes of the diameters associated with droplets of known concentration and of the diameters associated with droplets of the unknown concentration as much as possible about the unknown concentration. Such a direct comparison between the patient's specimen and the standard is usually necessary because the relationship between diameter and concentration usually varies from batch to batch of the medium.

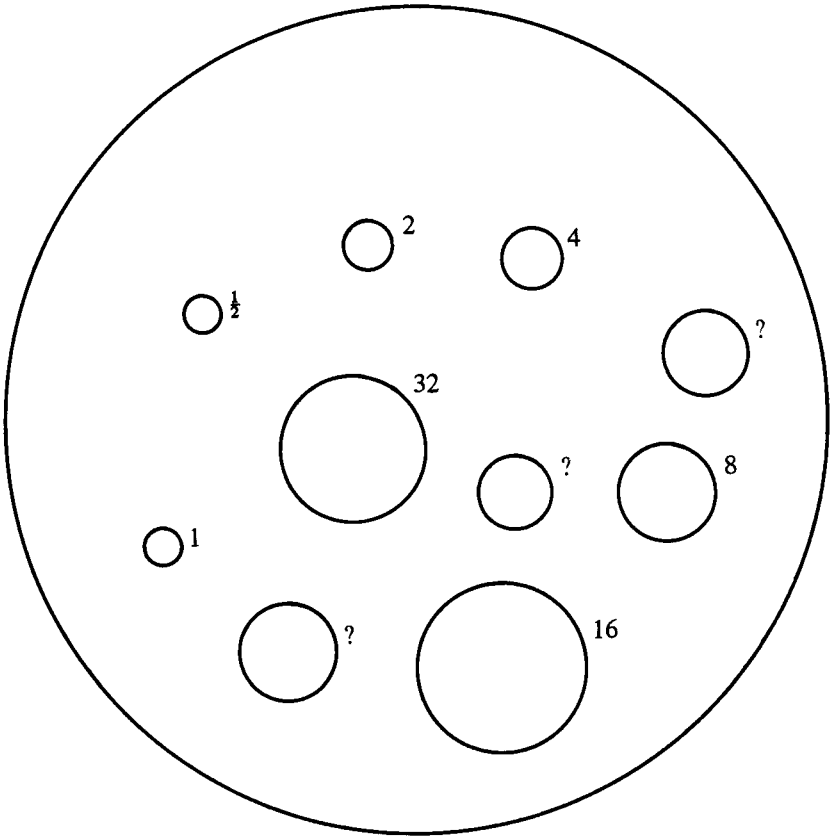

Fig. 1.1 Typical clearance circles in an antibiotic assay. For circles from standard droplets the known concentrations (mcg/ml) are shown. The three circles labelled ? are from droplets from a single specimen of unknown concentration.

Table 1.5 shows the results of an experiment to investigate the feasibility and reliability of this type of assay for a particular antibiotic. From the same batch of infected medium 20 Petri dishes were prepared and on each one droplet at each of seven concentrations (mcg/ml) was placed and the resulting clearance diameters recorded. Typical questions that we have to be in a position to answer are the following.

(1) Suppose that a droplet from a particular specimen has been placed on medium from this batch and has cleared a circle of diameter 19 mm. What can we infer about the concentration of antibiotic in the specimen?

(2) If three droplets from the same blood specimen have given clearance diameters of 18.0, 19.5 and 19.5 mm, what can be inferred about the concentration of the antibiotic in the specimen?

Table 1.5 *Clearance diameters (mm) on 20 Petri dishes for different concentrations (mcg/ml) of antibiotic*

Dish no.	Concentration (mcg/ml) 32	16	8	4	2	1	0.5
1	21.5	22.5	22.0	20.0	15.0	10.0	*
2	24.5	19.0	19.0	18.5	17.0	10.0	*
3	21.0	22.0	20.0	18.0	16.5	12.0	5.0
4	20.5	20.5	21.0	19.0	17.0	7.0	*
5	21.5	23.0	20.0	19.5	18.0	11.0	7.0
6	22.0	21.0	21.0	19.0	16.0	12.0	5.0
7	21.0	21.0	21.5	19.5	17.0	8.0	*
8	20.5	22.0	20.5	19.5	17.0	11.0	9.0
9	21.5	22.0	21.5	18.0	17.5	13.0	6.5
10	22.5	22.5	20.5	18.5	16.0	11.5	4.0
11	22.5	21.0	20.0	19.0	17.0	9.5	*
12	22.5	21.5	22.0	21.5	18.0	11.0	*
13	21.5	22.0	20.0	20.0	19.5	12.0	*
14	22.0	21.0	22.0	19.5	16.0	12.5	*
15	23.5	23.0	20.0	19.5	17.0	11.5	*
16	21.5	21.0	22.0	19.0	16.0	6.5	5.0
17	21.0	22.0	20.5	19.5	18.0	11.0	*
18	22.5	22.0	21.0	19.0	16.0	11.5	*
19	22.0	22.0	20.5	18.5	17.0	10.0	*
20	21.0	22.5	21.5	20.5	17.5	11.0	7.0

An entry * indicates that no measurable clearance was achieved.

(3) For medical reasons we may wish to be reasonably sure that the concentration quoted for a particular specimen is within 10 per cent of the true value. How many droplets from the specimen should be used to achieve this reliability?

Let f_t denote an experiment which records the clearance diameter of a droplet of concentration t of the antibiotic, and consider the class

$$F = \{f_t : t \in T\}$$

of experiments indexed by the set T of all possible concentrations. Then the informative experiment e which yields the data of table 1.5 is clearly of regression type and may be expressed briefly as

$$e = \{f_{t_1}, \ldots, f_{t_{140}}\},$$

a set of 140 independent experiments, where $t_1, \ldots, t_{140}$ are the concentrations associated with the 140 droplets. The data from this informative experiment thus consist of 140 pairs $(t_1, x_1), \ldots, (t_{140}, x_{140})$ of concentrations and corresponding clearance diameters, and can thus be set out in a typical scatter diagram (fig 1.2). For convenience a logarithmic concentration scale has been used.

In considering the first question posed we have the outcome $y = 19$ of some experiment from the class F, say f_u where u is the unknown concentration.

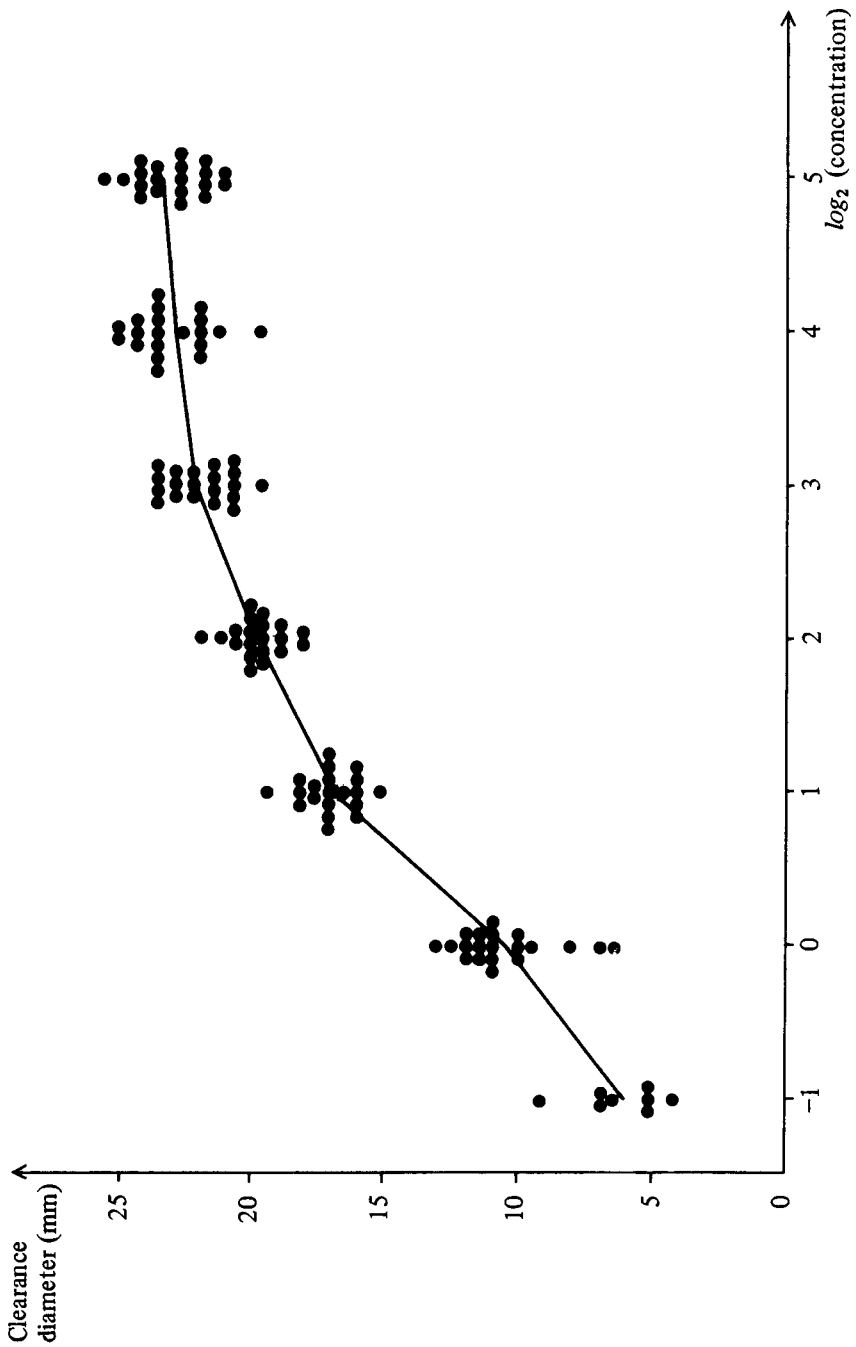

Fig. 1.2 Variability of clearance diameter with concentration of antibiotic. The line shown joins the mean clearance diameters at the different concentrations.

Table 1.6 *Results of eight preoperative tests on patients with Conn's syndrome*

		Concentrations in blood plasma					Blood Pressures	
Patient no.	Age (Years)	Na (meq/l)	K (meq/l)	CO_2 (meq/l)	Renin (meq/l)	Aldo-sterone (meq/l)	Systolic (mm Hg)	Diastolic (mm Hg)
*A*1	40	140.6	2.3	30.3	4.6	121.0	192	107
*A*2	37	143.0	3.1	27.1	4.5	15.0	230	150
*A*3	34	140.0	3.0	27.0	0.7	19.5	200	130
*A*4	48	146.0	2.8	33.0	3.3	30.0	213	125
*A*5	41	138.7	3.6	24.1	4.9	20.1	163	106
*A*6	22	143.7	3.1	28.0	4.2	33.0	190	130
*A*7	27	137.3	2.5	29.6	5.4	52.1	220	140
*A*8	18	141.0	2.5	30.0	2.5	50.2	210	135
*A*9	53	143.8	2.4	32.2	1.5	68.9	160	105
*A*10	54	144.6	2.9	29.5	3.0	144.7	213	135
*A*11	50	139.5	2.3	26.0	2.6	31.2	205	125
*A*12	44	144.0	2.2	33.7	3.9	65.1	263	133
*A*13	44	145.0	2.7	33.0	4.1	38.0	203	115
*A*14	66	140.2	3.1	29.1	4.7	43.1	195	115
*A*15	39	144.7	2.9	27.4	0.9	65.1	180	120
*A*16	46	139.0	3.1	31.4	2.8	192.7	228	133
*A*17	48	144.8	1.9	33.5	3.8	103.5	205	132
*A*18	38	145.7	3.7	27.4	2.8	42.6	203	117
*A*19	60	144.0	2.2	33.0	3.2	92.0	220	120
*A*20	44	143.5	2.7	27.5	3.6	74.5	210	114

(*continued*)