1

The Bayes linear approach

The subject of this book is the qualitative and quantitative analysis of our beliefs, with particular emphasis on the combination of beliefs and data in statistical analysis. In particular, we will cover:

(i) the importance of partial prior specifications for problems which are too complex to allow us to make meaningful full prior specifications;

(ii) simple ways to use our partial prior specifications to adjust our beliefs given observations;

(iii) interpretative and diagnostic tools that help us, first, to understand the implications of our collections of belief statements and, second, to make stringent comparisons between what we expect to observe and what we actually observe;

(iv) general approaches to statistical modelling based upon partial exchangeability judgements;

(v) partial graphical models to represent our beliefs, organize our computations and display the results of our analysis.

Our emphasis is methodological, so that we will mostly be concerned with types of specification and methods of analysis which are intended to be useful in a wide variety of familiar situations. In many of these situations, it will be clear that a careful, quantitative study of our beliefs may offer a valuable contribution to the problem at hand. In other cases, and in particular in certain types of problem that are conventionally treated by statisticians, the status of a belief analysis may be more controversial. Therefore, we shall begin our account by giving our views as to the role of the analysis of beliefs in such problems, and then briefly discuss what we perceive to be the strengths and weaknesses of the traditional Bayesian approach to belief analysis. We will briefly describe some of the distinctive features of Bayes

Bayes Linear Statistics: Theory and Methods M. Goldstein and D.A. Wooff

linear analysis, give an overview of the contents of this book and introduce the methodology by example.

1.1 Combining beliefs with data

To introduce our approach, compare the following examples. First, we test an individual for precognitive powers, and observe correct guesses in ten out of ten flips of a fair coin. Secondly, we test a promising new treatment against a current treatment for a disease, and observe that the new treatment outperforms the current treatment in each of ten trials on carefully matched pairs of patients.

The two experiments have, in a sense, yielded the same data, namely ten successes in ten binary trials. However, in the first case, most people would be intrigued but remain unconvinced that precognition had been demonstrated, whereas in the second case most people would be largely convinced of the efficacy of the new treatment. Such disagreements that might arise in the above analyses would be based, in the first case, on the extent of our predisposition to accept the existence of psychic powers, and, in the second, on possible medical grounds that we might have to be suspicious of the new treatment. Thus, similar data in different experiments may lead to different conclusions, when judged by the same person, and the same data may lead to different conclusions when judged by different people. In the above cases, the differences in the conclusions arise from differences in beliefs, either over the a priori plausibility of the hypotheses in the two experiments, or disagreements between individual beliefs as to the a priori plausibility of the hypothesis in a given experiment. More generally, people may disagree as to the relevance of the data to the conclusions or to any other feature of the probabilistic modelling required to reach a given conclusion.

Statistical theory has traditionally been concerned with analysing evidence derived from individual experiments, employing seemingly objective methods which lead to apparently clear-cut conclusions. In this view, the task of the statistician is to analyse individual data sets and, where necessary, pass the conclusions of the analysis to subject area specialists who then try to reach substantive conclusions. This viewpoint has the apparent virtue of turning statistics into a well-defined technical activity, which can be conducted in comparative isolation from the difficulties involved in making practical decisions. For example, in each of the two experiments above we may agree that, given a certain null hypothesis (no precognitive ability, no difference between treatments), the experiment has yielded a surprising result. This data analysis may be useful and revealing. However, as we have observed, such surprise may have different implications between experiments and between individuals. Ultimately, whether or not a particular data set suggests that a new treatment is better than the current treatment is only of interest if such consideration helps us to address the substantive question as to whether it is reasonable for us to believe and act as though the new treatment actually is better.

Such substantive analyses are much harder than the analysis of individual data sets, as they must confront and synthesize all of the evidence, including much that is fragmentary, contradictory, hard to find and difficult to assess, and for which there may be legitimate grounds for expert disagreement. However, these difficulties are unavoidable given that we want to reach substantive conclusions.

In practice, statisticians often do present themselves as addressing substantive issues, and are generally perceived as so doing by their clients. Indeed, the theory of statistical inference is generally formulated and perceived as an attempt to address substantive questions, but this may only be achieved within a traditional statistical analysis when the data set is sufficiently large and unambiguous as to overwhelm all other sources of prior information. When the statistical analysis is less clear-cut, it is necessary to synthesize the statistical results with all of the other considerations which might influence the substantive conclusions of the analysis. However, in current practice, this synthesis rarely takes place. As a result, the fate of far too many statistical analyses is to be accepted uncritically, or completely ignored, or treated in some other equally arbitrary fashion. The only way to avoid this fate is to frame the statistical analysis within the wider context with which the problem should be concerned, so that the purpose and construction of the analysis are directed at those things that we actually wish to know.

However, such a change in orientation requires a change in attitude and approach. Statisticians are used to being careful and precise in the collection and quantitative analysis of data. What we must further develop are the corresponding methods and skills for the specification and quantitative analysis of beliefs. As our beliefs are of fundamental interest, the study and refinement of these beliefs offer a central unifying principle for the bewildering variety of problems that we may confront when analysing uncertain situations.

The most fully developed methodology for such study is the Bayesian approach. We shall develop an alternative framework for the quantitative elicitation, analysis and interpretation of our beliefs, with particular emphasis on situations where our beliefs are at least partly influenced by statistical data. The framework is similar in spirit to the Bayes formalism. However, it differs in various important ways which are directed towards clearer and simpler analyses of beliefs, as, for reasons that we shall discuss in the next section, even the Bayesian approach can easily become, in practice, a methodology for using beliefs to analyse data, rather than a methodology for using data to analyse beliefs.

1.2 The Bayesian approach

Suppose that you visit a doctor, as you fear that you might have some particular disease, which you either have, event D, or you do not have, event D^c. The doctor gives you a test, which either is positive, event T, or not positive, event T^c. Before testing, you have a prior probability, $\mathrm{P}(D)$, that you have the disease. If you take the test, and the result is positive, then your conditional probability of the disease is given by Bayes' theorem as

$$\mathrm{P}(D|T) = \frac{\mathrm{P}(T|D)\mathrm{P}(D)}{\mathrm{P}(T|D)\mathrm{P}(D) + \mathrm{P}(T|D^c)\mathrm{P}(D^c)}. \tag{1.1}$$

Using Bayes' theorem, we replace the question

- Does the data, i.e. the test, suggest that you have the disease?

with the substantive question

- Should you now believe that you have the disease?

The evidence provided by the data, in this case the **likelihood ratio**,

$$\frac{\mathrm{P}(T|D)}{\mathrm{P}(T|D^c)},$$

has been combined with the external evidence as to whether you have the disease, as summarized by the **prior odds ratio**,

$$\frac{\mathrm{P}(D)}{\mathrm{P}(D^c)},$$

to produce the composite conditional probability $\mathrm{P}(D|T)$.

This form of argument dates back at least to the famous posthumously published essay of Thomas Bayes. At that time, probabilistic judgements were generally taken to be subjective quantifications of opinion. Subsequently, however, a different tradition arose, within which statisticians became reluctant to allow that a general statement, for example that a new treatment is better than a current treatment, could meaningfully be given a prior probability. As a result, use of the Bayes argument fell out of fashion, and probabilistic analysis was only deemed relevant within statistics to the extent that it applied to the outcomes of well-defined and repeatable sampling experiments.

While this may even now be a majority view, Bayes methods have recently grown again in popularity. This is partly due to the influence of decision analysis, in which the Bayes paradigm fits very naturally, and partly as a consequence of the critical re-examination of the logical, philosophical and practical basis of statistical procedures. The strengths of the Bayes approach are, first, that it appears to be more logical than most other approaches, replacing *ad hoc* methods with a unified methodology, and, second, that the approach may be used to address complex problems which cannot easily be considered within more traditional statistical paradigms. As a result, the approach has been judged to be successful in many applications, particularly where the analysis of data has been improved by combination with expert judgements.

However, perhaps because of the historical development, Bayes methods have themselves often been viewed as a sophisticated form of data analysis, so that much emphasis has been placed on ‘objective Bayes methods’ based on ‘non-informative priors’ and similar methods which are intended to extract information from a particular data set, without imposing any particular prior quantifications. Thus, there has developed a form of ‘objective’ Bayes methodology which is implicitly based around the idea that we may use beliefs to improve the analysis of data, in the sense that we may consider that data have a story to tell that is quite separate from the individual preconceptions that we may bring to the analysis. Such methods may be interesting, particularly for the analysis of large data sets, but they cannot address directly the substantive questions that concern us. To do so, we require the reverse process, namely to use data to help analyse beliefs. However, there is a fundamental difficulty in carrying out this program within the Bayes paradigm, namely that honest belief specification for large problems is usually very difficult.

Even in small problems, with few sources of uncertainty, it can be hard to distil all of our prior knowledge into a satisfactory full joint prior probability specification over all of the possible outcomes. In practical problems there may be hundreds of relevant sources of uncertainty about which we may make prior judgements. In such problems it is arguably impossible for us to carry out the Bayes programme, which requires us to specify meaningful probabilistic beliefs over collections of probability distributions over such high-dimensional structures. Even were we able to carry out such a full prior specification, we would usually find that the specification was too time-consuming and too difficult to check, document and validate to be worth the effort, unless we were working on questions that were of such importance that they justified the enormous expenditure of effort that is required simply to apply the paradigm in an honest fashion.

Even if we were able to make such high-dimensional specifications, the resulting Bayes analysis would often be extremely computer intensive, particularly in areas such as experimental design. Computational issues, while of great practical importance, are secondary to the fundamental difficulty of making meaningful high-dimensional prior probability specifications. However, such considerations do support the basic argument that we shall develop in this book, which is as follows.

The more complex the problem, the more we need help to consider the resulting uncertainties, but the more difficult it is to carry out a full Bayes analysis. Essentially, the Bayes approach falls victim to the ambition in its formulation. Often, the approach is considered to be a description of what a perfectly rational individual would do when confronted with the problem. The implication is that we should copy the behaviour of such an individual as closely as we can. However, as the complexity of problems increases, the disparity between the hypothetical abilities of the perfectly rational analyst and our actual abilities to specify and analyse our uncertainties becomes so wide that it is hard to justify the logical or practical relevance of such a formulation.

Therefore if, in complex problems, we are unable to make and analyse full prior specifications, it follows that we need to develop methods based around **partial** belief specification. We shall develop one such methodology, termed the **Bayes**

linear approach. The approach is similar in spirit to the full Bayes approach, and is particularly appropriate whenever the full Bayes approach requires an unnecessarily exhaustive description and analysis of prior uncertainty.

Depending on our viewpoint, we may view the Bayes linear approach either as offering a simple approximation to a full Bayes analysis, for problems where the full analysis would be too difficult or time-consuming, or as complementary to the full Bayes analysis, offering a variety of new interpretative and diagnostic tools which may be of value whatever our viewpoint, or as a generalization of the full Bayes approach, where we lift the artificial constraint that we require full probabilistic prior specification before we may learn anything from data.

1.3 Features of the Bayes linear approach

The following are important features of the Bayes linear approach.

1. The approach is subjectivist. We express our prior judgements of uncertainty in quantitative form, and adjust these uncertainties in the light of observation.

2. We use prior specifications which honestly correspond to prior beliefs. In order to do this, we must structure our analyses so that the prior specifications that we require are within the ability of the individual to make.

3. The approach is based on expectation rather than probability as a primitive. With expectation as a primitive, we may immediately obtain probabilities as expectations of indicator functions. With probability as a primitive, we need to determine all probabilities for a quantity before we may assess the expectation. Therefore, starting with expectation allows us to focus directly on the crucial uncertainties in the problem.

4. With expectation as a primitive, the fundamental object of interest is the collection of random quantities, which are naturally gathered into inner product spaces. Therefore, the resulting analysis follows from the geometric structure implied by the partial belief specification.

5. Beliefs are adjusted by linear fitting rather than conditioning. Therefore, the Bayes linear approach may be viewed as a simple and tractable approximation to a full Bayes analysis.

6. There are general temporal relationships between the adjusted beliefs created by linear fitting and our posterior beliefs. Full conditioning is a special case of linear fitting whose general temporal relation with posterior beliefs is no different than for any other linear fit. Therefore the full Bayes analysis may be also viewed as a particular special case of the Bayes linear approach.

7. As linear fitting is generally computationally simpler than full conditioning, we may often analyse complex problems, in particular those arising in experimental design, more straightforwardly than under the full Bayes counterpart.

8. We only specify beliefs over observable quantities, so that all of our belief statements can be given a direct, physical interpretation. We therefore construct underlying population models strictly by means of exchangeability judgements over observables, which is feasible precisely because we take expectation as the primitive for the theory.

9. Our aim is to develop improved assessments of belief. Partly, this is achieved by sensible processing of prior and data inputs. However, just as important is the qualitative interpretation of the belief adjustment. Therefore, we develop interpretative tools to identify which aspects of our prior judgements and the data are most influential for which aspects of our conclusions, so that we may judge whether or not our belief adjustments appear intuitively reasonable, and compare possible alternative adjustments, based for example on different sampling frames or experimental designs.

10. When we adjust our beliefs, we similarly need qualitative methods for interpreting the resulting collection of changes in belief. Therefore, we develop interpretative tools to summarize both the magnitude and the nature of the overall changes in belief, and to display conflict or consistency between the various sources of evidence which contribute to such changes.

11. Each belief statement made about an observable may be subsequently compared with the value of that observable. Stringent diagnostics are available to warn us of possible conflicts between our beliefs and reality.

12. There are important special cases, for example certain analyses for multivariate Gaussian models, where many aspects of the Bayes and the Bayes linear approaches correspond. Therefore, many of the interpretative and diagnostic tools that we describe will also be relevant for such analyses. Further, it is of general interest to separate those aspects of the Gaussian analysis which follow directly from the geometric implications of the second-order specification, from those aspects whose validity depends on the precise form of the Gaussian density function.

13. Much of the qualitative and quantitative structure of the Bayes linear analysis may be displayed visually using Bayes linear graphical models. These models aid the intuitive understanding of expected and observed information flow through complex systems, and also facilitate efficient local computation methods for the analysis of large systems.

1.4 Example

As a trailer for the ideas in the book we give the following example. The example is intended to convey the flavour of our approach, and so we refrain both from detailed exposition of the methodology and from deep analysis of the problem.

A factory produces two products. For planning purposes, the factory wishes to predict sales of the products in each period. In order to do this, various relevant information will be used, in particular the sales of the two products in the previous period. For this introduction, it will be sufficient to suppose that this is all that is explicitly used, though of course the judgements of the sales forecasters will be called on to formulate the prior beliefs.

For illustration, we shall imagine that sales at a time point soon to come are used to improve our understanding of sales at a more distant future time point. Thus, there are four quantities of interest: X_1 and X_2, the sales of products 1 and 2 at the first time point, and Y_1 and Y_2, the corresponding sales at the later time point.

For the simplest form of analysis that we shall describe, the sales forecaster first specifies prior expectations for the four quantities, together with a variance matrix over them. We will consider the problem of eliciting and specifying prior information in the form of expectations, variances, and covariances in Chapter 2. In the meantime, suppose that we have based our prior specifications on sample information from previous sales figures, and managerial judgements as to their relevance in the light of any special circumstances which may be felt appropriate to the current sales period.

1.4.1 Expectation, variance, and standardization

In this book, we assume basic knowledge of expectation, variance and covariance, and correlation. Suppose that X and Y are collections of m and n random quantities, respectively. The expectation for X is denoted by $\mathrm{E}(X)$, an $m \times 1$ vector with ith element $\mathrm{E}(X_i)$. The variance for X is denoted by $\mathrm{Var}(X)$, an $m \times m$ variance–covariance matrix with (i, i)th element $\mathrm{Var}(X_i)$ and with (i, j)th element giving the covariance between X_i and X_j, denoted by $\mathrm{Cov}(X_i, X_j)$. The covariance between X and Y is denoted by $\mathrm{Cov}(X, Y)$, an $m \times n$ covariance matrix with (i, j)th element $\mathrm{Cov}(X_i, Y_j)$. The correlation between X and Y is denoted by $\mathrm{Corr}(X, Y)$, an $m \times n$ correlation matrix with (i, j)th element $\mathrm{Corr}(X_i, Y_j)$, assuming finite non-zero variances $\mathrm{Var}(X_i)$ and $\mathrm{Var}(Y_j)$. We may find it helpful to refer to the standardized versions of quantities.

Definition 1.1 *For a random quantity X, we write the standardized quantity as*

$$\mathrm{S}(X) = \frac{X - \mathrm{E}(X)}{\sqrt{\mathrm{Var}(X)}}.$$

1.4.2 Prior inputs

Suppose that, in some appropriate units, the prior mean for each quantity is 100; the prior variance for X_1, X_2 is 25; the prior variance for the future sales Y_1, Y_2 is 100; and the prior correlation matrix over all four quantities is

	X_1	X_2	Y_1	Y_2
X_1	1.00	−0.60	0.60	−0.20
X_2	−0.60	1.00	−0.20	0.60
Y_1	0.60	−0.20	1.00	−0.60
Y_2	−0.20	0.60	−0.60	1.00

Thus, we might summarize our prior specifications as follows. We have the same expectation for sales for each product at each time point, but we are much less certain about the sales figures for the later time point. The correlation matrix specified expresses the belief that sales of each product are quite strongly positively correlated over the two time periods, but that the products are considered to compete and so sales of the two products are negatively correlated. Note that in this problem we do not complete the prior specification by choosing a prior joint probability distribution for these four quantities with the given mean and variance structure. Rather, our aim is to perform an analysis based solely on the partial prior specification that we have described.

We intend to use the sales at the first time point to improve our forecasts for sales at the later time point. Much of our approach deals with simultaneous analysis of **collections** of quantities, so, for convenience, we group together the two sales from the first time point into the collection $D = (X_1, X_2)$, and the two sales for the later time point into the collection $B = (Y_1, Y_2)$. There is no particular significance to the names B and D, except that we sometimes find it useful to retain D for a collection of 'data' quantities (i.e. quantities which we intend to observe, and so for which data will become available) and to retain B for a collection of 'belief' quantities (i.e. quantities that we wish to predict, and so for which we have prior beliefs followed by adjusted beliefs).

1.4.3 Adjusted expectations

There are many ways in which we might try to improve our forecasts for the collection B. A simple method, which exploits the prior mean and variance statements that we have made, is as follows. We can look among the collection of linear estimates, i.e. those of the form $c_0 + c_1X_1 + c_2X_2$, and choose constants c_0, c_1, c_2 to minimize the prior expected squared error loss in estimating each of Y_1 and Y_2. For example, we aim to minimize

$$\mathrm{E}([Y_1 - c_0 - c_1X_1 - c_2X_2]^2). \tag{1.2}$$

The choices of constants may be easily computed from the above specifications, and the estimators turn out to be

$$\mathrm{E}_D(Y_1) = 1.5X_1 + 0.5X_2 - 100, \tag{1.3}$$

$$\mathrm{E}_D(Y_2) = 0.5X_1 + 1.5X_2 - 100. \tag{1.4}$$

We call $\mathrm{E}_D(Y_1)$ the **adjusted expectation** for Y_1 given the information $D = [X_1, X_2]$. Similarly, $\mathrm{E}_D(Y_2)$ is the adjusted expectation for Y_2 given D. The adjusted expectations have a number of properties which we will come to below and in later chapters; in particular, they are themselves random quantities, so that they too have expectations, variances and so forth.

1.4.4 Adjusted versions

We will be concerned not only with the adjusted expectation for a quantity, but also with the residual component associated with it, which we call the **adjusted version** of the quantity. The adjusted version of Y given D is defined to be $\mathbb{A}_D(Y) = Y - \mathrm{E}_D(Y)$. In our example, the adjusted versions are

$$\mathbb{A}_D(Y_1) = Y_1 - (1.5X_1 + 0.5X_2 - 100), \tag{1.5}$$

$$\mathbb{A}_D(Y_2) = Y_2 - (0.5X_1 + 1.5X_2 - 100). \tag{1.6}$$

These adjusted versions have important roles to play in Bayes linear analysis in that they allow us to quantify the uncertainty expected to remain after an adjustment. A priori, we expect the residual component to be zero, $\mathrm{E}(\mathbb{A}_D(Y_i)) = 0$.

1.4.5 Adjusted variances

How useful are the adjusted expectations when judged as predictors? One way to assess how much information about the elements of B we gain by observing the elements of D is to evaluate the **adjusted variance** for each quantity. The adjusted variance for any quantity Y, given a collection of information D, is defined as

$$\mathrm{Var}_D(Y) = \mathrm{Var}(\mathbb{A}_D(Y)) = \mathrm{E}([Y - \mathrm{E}_D(Y)]^2),$$

being the minimum of the prior expected squared error loss in the sense of (1.2). This is a measure of the residual uncertainty, or, informally, the 'unexplained' variance, having taken into account the information in D. The portion of variation resolved is

$$\mathrm{Var}(Y) - \mathrm{Var}_D(Y) = \mathrm{Var}(\mathrm{E}_D(Y)).$$

For this example the adjusted variances are the same, so that we have

$$\mathrm{Var}_D(Y_1) = \mathrm{Var}_D(Y_2) = 60,$$

whereas we began with variances $\mathrm{Var}(Y_1) = \mathrm{Var}(Y_2) = 100$. Consequently, the value of observing sales at the first time point is to reduce our uncertainty about sales at the later time point by 40%. We typically summarize the informativeness of data D for any quantity Y by a scale-free measure which we call the **resolution** of Y induced by D, defined as

$$\mathrm{R}_D(Y) = 1 - \frac{\mathrm{Var}_D(Y)}{\mathrm{Var}(Y)} = \frac{\mathrm{Var}(\mathrm{E}_D(Y))}{\mathrm{Var}(Y)}. \tag{1.7}$$

In our example, the variance resolutions are $\mathrm{R}_D(Y_1) = \mathrm{R}_D(Y_2) = 0.4$. The resolution lies between 0 and 1, and in general, small (large) resolutions imply that the information has little (much) linear predictive value, given the prior specification.

In terms of the vector B, we began with a variance matrix $\mathrm{Var}(B)$ which we have decomposed into unresolved and resolved portions, each a matrix:

$$\mathrm{Var}(B) = \mathrm{Var}_D(B) + \mathrm{RVar}_D(B), \tag{1.8}$$

where $\mathrm{RVar}_D(B) = \mathrm{Var}(\mathrm{E}_D(B))$ is our notation for the resolved variance matrix for the adjustment of the collection B by the collection D, and equals the prior variance matrix for the adjusted expectation vector. The off-diagonal terms are **adjusted covariances** and **resolved covariances**. For example, the adjusted covariance between Y_1 and Y_2 given D is the covariance between the two residual components,

$$\mathrm{Cov}_D(Y_1, Y_2) = \mathrm{Cov}(\mathbb{A}_D(Y_1), \mathbb{A}_D(Y_2)),$$

and the resolved covariance is the change from prior to adjusted,

$$\mathrm{RCov}_D(Y_1, Y_2) = \mathrm{Cov}(Y_1, Y_2) - \mathrm{Cov}_D(Y_1, Y_2).$$

In our example, the decomposition (1.8) turns out to be

$$\mathrm{Var}(B) = \begin{bmatrix} 100 & -60 \\ -60 & 100 \end{bmatrix} = \begin{bmatrix} 60 & -60 \\ -60 & 60 \end{bmatrix} + \begin{bmatrix} 40 & 0 \\ 0 & 40 \end{bmatrix}.$$

The off-diagonal entries here show that $\mathrm{Cov}(Y_1, Y_2) = \mathrm{Cov}_D(Y_1, Y_2) = -60$, and that $\mathrm{RCov}_D(Y_1, Y_2) = 0$. It may seem a little puzzling that we do not seem to have resolved any of the covariance between Y_1 and Y_2. Indeed, the variance matrix for their adjusted versions is singular. We shall discover why this is so, and comment on it in more detail, later.

1.4.6 Checking data inputs

At some point, we may observe the values of D. In our case, suppose that the sales at the first time point turn out to be $x_1 = 109$ and $x_2 = 90.5$. (We follow convention in using lower case for observations and upper case for unknowns.) The first thing we do is to check that these observations are consistent with beliefs specified about them beforehand. A simple diagnostic is to examine the **standardized change** from the prior expectation to the observed value. In our example, the standardized changes are

$$\mathrm{S}(x_1) = \frac{x_1 - \mathrm{E}(X_1)}{\sqrt{\mathrm{Var}(X_1)}} = \frac{109 - 100}{\sqrt{25}} = 1.8, \tag{1.9}$$

$$\mathrm{S}(x_2) = \frac{90.5 - 100}{\sqrt{25}} = -1.9. \tag{1.10}$$

Each (squared) standardized change has prior expectation one. Informally, we might begin to suspect an inconsistency if we saw a standardized change of more than about two standard deviations; and be quite concerned to see standardized changes of more than about three standard deviations. We do not wish to give rigid rules or thresholds for interpreting these kinds of measure, as they are largely dependent on the context of the problem.

1.4.7 Observed adjusted expectations

When the data quantities are observed we may calculate the observed adjusted expectations. Replacing X_1, X_2 by $x_1 = 109$ and $x_2 = 90.5$ in (1.3) and (1.4), we obtain the following assessments:

$$\mathrm{E}_d(Y_1) = 1.5 \times 109 + 0.5 \times 90.5 - 100 = 108.75,$$

$$\mathrm{E}_d(Y_2) = 0.5 \times 109 + 1.5 \times 90.5 - 100 = 90.25.$$

We call these values **observed adjusted expectations**. Notice that our subscript notation uses lower case, $\mathrm{E}_d(\cdot)$, rather than upper case, $\mathrm{E}_D(\cdot)$ to indicate that the entire collection D has been observed to be d. The effect of the data here is to cause our expectations for future sales to follow a similar pattern, i.e. larger and smaller sales respectively in the two components.

1.4.8 Diagnostics for adjusted beliefs

It is valuable at this stage to check how different the observed adjusted expectation is from the prior expectation. A simple diagnostic is given by the change from prior to adjusted expectation, standardized with respect to the variance of the adjusted expectation. We have that $\mathrm{E}(\mathrm{E}_D(Y)) = \mathrm{E}(Y)$ for any Y and D. Thus, from (1.9), the standardized change is

$$\mathrm{S}(\mathrm{E}_d(Y)) = \frac{\mathrm{E}_d(Y) - \mathrm{E}(Y)}{\sqrt{\mathrm{Var}(\mathrm{E}_D(Y))}},$$

where the denominator in the standardization does not depend on the observed data. We call these standardized changes the **standardized adjustments**. In our example, they are:

$$\mathrm{S}(\mathrm{E}_d(Y_1)) = \frac{108.75 - 100}{\sqrt{40}} = 1.38, \qquad \mathrm{S}(\mathrm{E}_d(Y_2)) = \frac{90.25 - 100}{\sqrt{40}} = -1.54,$$

where in each case the squared standardized adjustment has prior expectation one. As such, the changes in expectation for sales at a future time point are 1.38 and 1.54 standard deviations, relative to variation explained, and so are roughly in line with what we expected beforehand.

1.4.9 Further diagnostics for the adjusted versions

As time progresses, we eventually discover actual sales, $y_1 = 112$ and $y_2 = 95.5$, of the two products. It is diagnostically important now to compare our predictions with what actually happened. There are two diagnostics to examine. First, we can compare a quantity's observation with its prior expectation, irrespective of the linear fitting on D. The standardized change in expectation for a quantity is given by (1.9). In our example, the standardized changes in expectation from prior to observed are $\mathrm{S}(Y_1) = (112 - 100)/10 = 1.2$ and $\mathrm{S}(Y_2) = -0.45$, so these future sales turned out to be consistent with our prior considerations.

A second diagnostic is given by examining the change from adjusted expectation to actual observation, relative to the associated adjusted variance, as this was the variation remaining in each Y_i after fitting on D, but before observing Y_1 and Y_2. By observing the actual sales values y_1, y_2, we observe the residual components, i.e. the adjusted versions $\mathbb{A}_D(Y_i) = Y_i - \mathrm{E}_D(Y_i)$. Given that they had prior expectation zero, we wish to see how far the adjusted versions have changed from zero, relative to their variances

$$\mathrm{Var}(\mathbb{A}_D(Y_i)) = \mathrm{Var}_D(Y_i).$$

The appropriate standardized change is thus

$$\mathrm{S}_d(y_i) = \mathrm{S}(\mathbb{A}_d(y_i)) = \frac{y_i - \mathrm{E}_d(Y_i)}{\sqrt{\mathrm{Var}_D(Y_i)}}.$$

In our example, the sales at the later time point, $y_1 = 112, y_2 = 95.5$, should be compared to the adjusted expectations $\mathrm{E}_d(Y_1) = 108.75$ and $\mathrm{E}_d(Y_2) = 90.25$, standardizing with respect to the adjusted variances:

$$\mathrm{Var}_D(Y_1) = \mathrm{Var}_D(Y_2) = 60.$$

We obtain

$$\mathrm{S}_d(y_1) = \frac{112 - 108.75}{\sqrt{60}} = 0.42 \quad \text{and} \quad \mathrm{S}_d(y_2) = \frac{95.5 - 90.25}{\sqrt{60}} = 0.68.$$

The squared standardized changes should again be about one, so our diagnostic checks suggest that both of our predictions were roughly within the tolerances suggested by our prior variance specifications. If anything, the adjusted expectations are, in terms of standard deviations, rather closer to the observed values than expected.

1.4.10 Summary of basic adjustment

Let us summarize our results so far in the form of tables, shown in Table 1.1. The analysis results in decomposing the sales quantities into two parts, the first of which comes from linear fitting on other quantities D, and the second of which is residual. Summary statistics are calculated for the original and component quantities; all summaries are additive over components, except for the standardized changes. We note that the diagnostics reveal nothing untoward: all the standardized changes are about in line with what was expected beforehand. In each case, the change from prior to adjusted expectation was slightly larger than expected, one up and one down; and in each case the standardized change from adjusted expectation to

Table 1.1 Adjusting future sales Y_1, Y_2 by previous sales: summary.

	Original	=	Adjusted expectation	+	Adjusted version
Quantity	Y_1	=	$\mathrm{E}_D(Y_1)$	+	$\mathbb{A}_D(Y_1)$
		=	$1.5X_1 + 0.5X_2 - 100$	+	$Y_1 - \mathrm{E}_D(Y_1)$
Prior	$\mathrm{E}(Y_1)$	=	$\mathrm{E}(\mathrm{E}_D(Y_1)) = \mathrm{E}(Y_1)$	+	$\mathrm{E}(\mathbb{A}_D(Y_1)) = 0$
expectation	100	=	100	+	0
Prior	$\mathrm{Var}(Y_1)$	=	$\mathrm{RVar}_D(Y_1)$	+	$\mathrm{Var}_D(Y_1)$
variance	100	=	40	+	60
Observed	y_1	=	$1.5x_1 + 0.5x_2 - 100$	+	$y_1 - \mathrm{E}_d(Y_1)$
	112	=	108.75	+	3.25
Standardized	$\frac{y_1 - \mathrm{E}(Y_1)}{\sqrt{\mathrm{Var}(Y_1)}}$		$\frac{\mathrm{E}_d(Y_1) - \mathrm{E}(Y_1)}{\sqrt{\mathrm{RVar}_D(Y_1)}}$		$\frac{y_1 - \mathrm{E}_d(Y_1)}{\sqrt{\mathrm{Var}_D(Y_1)}}$
change	1.2		1.38		0.42
Quantity	Y_2	=	$\mathrm{E}_D(Y_2)$	+	$\mathbb{A}_D(Y_2)$
		=	$0.5X_1 + 1.5X_2 - 100$	+	$Y_2 - \mathrm{E}_D(Y_2)$
Prior	$\mathrm{E}(Y_2)$	=	$\mathrm{E}(\mathrm{E}_D(Y_2)) = \mathrm{E}(Y_2)$	+	$\mathrm{E}(\mathbb{A}_D(Y_2)) = 0$
expectation	100	=	100	+	0
Prior	$\mathrm{Var}(Y_2)$	=	$\mathrm{RVar}_D(Y_2)$	+	$\mathrm{Var}_D(Y_2)$
variance	100	=	40	+	60
Observed	y_2	=	$0.5x_2 + 1.5x_2 - 100$	+	$y_2 - \mathrm{E}_d(Y_2)$
	95.5	=	90.25	+	5.25
Standardized	$\frac{y_2 - \mathrm{E}(Y_2)}{\sqrt{\mathrm{Var}(Y_2)}}$		$\frac{\mathrm{E}_d(Y_2) - \mathrm{E}(Y_2)}{\sqrt{\mathrm{RVar}_D(Y_2)}}$		$\frac{y_2 - \mathrm{E}_d(Y_2)}{\sqrt{\mathrm{Var}_D(Y_2)}}$
change	−0.45		−1.54		0.68

observed value was smaller than expected, and closer to the original prior expectation. Whether this should cause concern cannot be answered solely by examining single quantities using summaries such as these, useful though they are. In fact, we need also to analyse changes in our collection of beliefs, which we consider next.

1.4.11 Diagnostics for collections

We showed in §1.4.6 how we check individual data inputs by calculating standardized changes. To check a collection of data inputs, we need to make a basic consistency check, and if this is successful we proceed to calculate a global discrepancy. For the basic consistency check, recall that, for any random quantity X, if we specify $\mathrm{Var}(X) = 0$ then we expect to observe $x = \mathrm{E}(X)$: otherwise either the variance specification is wrong, or perhaps some error has occurred in collecting the data. For a collection (vector) of random quantities B, with observed value b, expectation $\mathrm{E}(B)$, and variance matrix $\mathrm{Var}(B)$, the basic consistency check is as follows. If $\mathrm{Var}(B)$ is non-singular then the value of $b - \mathrm{E}(B)$ is unconstrained, and the basic consistency check is passed. Otherwise, $\mathrm{Var}(B)$ has one or more eigenvalues equal to zero. In this case, suppose that q is an eigenvector corresponding to a zero eigenvalue. Such eigenvectors identify linear combinations of the Bs having variance zero, as for each such eigenvector q, it is the case that $\mathrm{Var}(q^T B) = 0$. Consequently, in the case of singularity the basic consistency check lies in verifying that $q^T b = q^T \mathrm{E}(B)$ for every eigenvector q corresponding to a zero eigenvalue. Failure of the consistency check always corresponds to infinite values for the corresponding standardized changes. Following a successful basic consistency check, we calculate measures of discrepancy based on the Mahalanobis distance.

To return to checking data inputs, we are concerned with differences between a vector of data d and the vector of prior expectations $\mathrm{E}(D)$. The variance matrix concerned here is

$$\mathrm{Var}(D) = \begin{bmatrix} 25 & -15 \\ -15 & 25 \end{bmatrix},$$

which is full rank, so that the basic consistency check is passed. Next, for our measure of the difference between the data d and their prior expectations $\mathrm{E}(D)$, we calculate the **discrepancy**, $\mathrm{Dis}(d)$, as the Mahalanobis distance between d and $\mathrm{E}(D)$:

$$\begin{aligned} \mathrm{Dis}(d) &= (d - \mathrm{E}(D))^T \mathrm{Var}(D)^{\dagger} (d - \mathrm{E}(D)) \\ &= \begin{bmatrix} 109 - 100 & 90.5 - 100 \end{bmatrix} \begin{bmatrix} 25 & -15 \\ -15 & 25 \end{bmatrix}^{-1} \begin{bmatrix} 109 - 100 \\ 90.5 - 100 \end{bmatrix} \\ &= 4.29. \end{aligned}$$

Here, $\mathrm{Var}(D)^{\dagger}$ is the Moore–Penrose generalized inverse of $\mathrm{Var}(D)$, equivalent to the usual inverse $\mathrm{Var}(D)^{-1}$ when $\mathrm{Var}(D)$ is full rank. The Moore–Penrose inverse is employed as we make no distinction between the handling of full rank and singular variance matrices: this is especially useful when analysing the structural implications of prior specifications. The discrepancy has prior expectation equal to the rank of the prior variance matrix $\mathrm{Var}(D)$, which in our example has rank two. We thus obtain as a summary statistic of the discrepancy between the observed values and the prior specification, the **discrepancy ratio**,

$$\mathrm{Dr}(d) = \frac{\mathrm{Dis}(d)}{\mathbf{rk}\{\mathrm{Var}(D)\}} = 2.15,$$

to be compared to its prior expectation of one. For single observations rather than collections, the discrepancies are just the squared standardized changes. None of these measures indicate any substantial problem with our prior formulation.

We showed in §1.4.8 how we calculate a standardized adjustment to check for a difference between an observed adjusted expectation and the corresponding prior expectation. As above, we obtain a global diagnostic by making a basic consistency check and then calculating a measure of discrepancy. The vectors to be compared are the observed adjustments, $\mathrm{E}_d(B)$, and their prior expectations, $\mathrm{E}(B)$. The variance matrix concerned is

$$\mathrm{Var}(\mathrm{E}_D(B)) = \mathrm{RVar}_D(B) = \begin{bmatrix} 40 & 0 \\ 0 & 40 \end{bmatrix},$$

which is full rank, so that the basic consistency check is passed. We obtain a global diagnostic for the observed adjustment by calculating the Mahalanobis distance between the observed adjusted expectations and the prior expectations, to give the **adjustment discrepancy**, $\mathrm{Dis}_d(B)$, where

$$\begin{aligned} \mathrm{Dis}_d(B) &= (\mathrm{E}_d(B) - \mathrm{E}(B))^T \mathrm{RVar}_D(B)^{\dagger} (\mathrm{E}_d(B) - \mathrm{E}(B)) \\ &= \begin{bmatrix} 108.75 - 100 & 90.25 - 100 \end{bmatrix} \begin{bmatrix} 40 & 0 \\ 0 & 40 \end{bmatrix}^{-1} \begin{bmatrix} 108.75 - 100 \\ 90.25 - 100 \end{bmatrix} \\ &= 4.29. \end{aligned}$$

As before, this discrepancy measure fails to suggest any substantial problem with our prior formulation.

For our final collection diagnostic of this section, we showed in §1.4.9 how to calculate the standardized change from observed adjusted expectation, $\mathrm{E}_d(Y_i)$, to actual observation y_i, where the standardization is with respect to the variance remaining in Y_i, $\mathrm{Var}_D(Y_i)$, before observing it. As above, we proceed to a global diagnostic where we wish to measure the discrepancy between the observed adjusted expectations $\mathrm{E}_d(B)$, and the actual observations $b = [y_1 \ y_2]^T$, relative to the variance matrix $\mathrm{Var}_D(B)$. Another way of thinking about this is that we finally observe the adjusted versions $\mathbb{A}_D(B)$ and wish to see whether these observations are consistent with their prior variance–covariance specifications, $\mathrm{Var}(\mathbb{A}_D(B))$. For a basic consistency check, we have that

$$\mathrm{Var}(\mathbb{A}_D(B)) = \mathrm{Var}_D(B) = \begin{bmatrix} 60 & -60 \\ -60 & 60 \end{bmatrix}, \tag{1.11}$$

which is singular. There is one eigenvalue equal to zero, with corresponding eigenvector proportional to $[1 \ 1]^T$. Consequently we have specified a variance of zero for

$$\begin{bmatrix} 1 & 1 \end{bmatrix}^T \begin{bmatrix} \mathbb{A}_D(Y_1) \\ \mathbb{A}_D(Y_2) \end{bmatrix} = \mathbb{A}_D(Y_1) + \mathbb{A}_D(Y_2),$$

and it is thus necessary to verify in this example that the observed adjusted versions sum to their expected value, which is zero. However, we see from Table 1.1 that the observed adjusted versions are 3.25 and 5.25, summing to $8.5 \neq 0$, so we have discovered a very serious flaw in our specification. In practice there is no point in proceeding further with the analysis. Had the basic consistency check not failed, we would have calculated the adjusted version discrepancy as

$$\begin{aligned}&(b - \mathrm{E}_d(B))^T \mathrm{Var}_D(B)^{\dagger}(b - \mathrm{E}_d(B))\\ &= \begin{bmatrix}112 - 108.75 & 95.5 - 90.25\end{bmatrix}\begin{bmatrix}60 & -60\\ -60 & 60\end{bmatrix}^{\dagger}\begin{bmatrix}112 - 108.75\\ 95.5 - 90.25\end{bmatrix} = 0.02.\end{aligned}$$

1.4.12 Exploring collections of beliefs via canonical structure

To this point we have specified prior information, recorded some data, obtained predictions, calculated the value of the predictions, and compared expected to actual behaviour, largely focusing on the single quantities of interest, Y_1 and Y_2, the sales for two products at a future time point. Little of the analysis turned up anything surprising: changes in expectation were mostly about in line with what we expected. However, one of the diagnostics calculated for a collection revealed a very serious flaw, namely actual observations which should not have been possible given the prior specifications. This suggests, rightly, that our analysis should focus on analysing collections of beliefs, rather than on piecemeal analysis for single quantities. Further, to focus on collections of beliefs will allow us naturally to address many other relevant questions. For example, it reveals the implications of correlations between the collections of interest; it allows us to make global uncertainty and diagnostic assessments for entire collections or any sub-collections we choose; and it allows us easily to go beyond analysis of single quantities such as Y_1 and Y_2 to such quantities as total sales, $Y_1 + Y_2$, or the difference between sales, $Y_1 - Y_2$. Answering such questions is an important part of the Bayes linear approach.

It turns out, whether our interest is in making assessments for simple quantities such as Y_1, or for interesting linear combinations such as $Y_1 + Y_2$, or for global collections such as $B = [Y_1, Y_2]$, that for all such problems there is a natural reorganization which we may use to answer these questions directly. The reorganization arises by generating and exploiting an underlying **canonical structure**. This structure completely summarizes the global dynamics of belief adjustment for an analysis. For the two-dimensional problem, this amounts to finding the linear combinations of Y_1 and Y_2 about which D is respectively most and least informative, in the sense of maximizing and minimizing the variance resolution. In our example, these linear combinations have a particularly simple form; they are Z_1 and Z_2, where

$$Z_1 = 0.112(Y_1 + Y_2) - 22.361, \tag{1.12}$$

$$Z_2 = 0.056(Y_1 - Y_2). \tag{1.13}$$

For convenience, we have centred each Z_i so that it has prior mean zero, and scaled it so that it has prior variance one. We call Z_1 and Z_2 respectively the first and second **canonical directions**. Canonical directions are always uncorrelated. For our example, Z_1 is essentially a linear combination giving total sales, and Z_2 is the difference between sales. As far as the original sales quantities are concerned, they can be expressed in terms of the canonical quantities as

$$Y_1 = 4.472(Z_1 + 2Z_2) + 100,$$

$$Y_2 = 4.472(Z_1 - 2Z_2) + 100.$$

In addition to calculating the canonical directions, we also calculate their resolutions $\mathrm{R}_D(Z_1)$ and $\mathrm{R}_D(Z_2)$ from (1.7). We call these the **canonical resolutions**. The canonical directions and canonical resolutions together comprise the canonical structure. In our example, the resolutions in the canonical directions are $\mathrm{R}_D(Z_1) = 1$ and $\mathrm{R}_D(Z_2) = 0.25$. In the latter case, the implication is that the minimum variance resolution for **any** linear combination of the two unknown sales quantities is 0.25, i.e. by observing D we expect to 'explain' at least 25% of the variance for **all** linear combinations of our future sales quantities, Y_1 and Y_2.

The resolution of Z_1 turns out to be exactly 1. This means that, according to our prior specifications, there will be no uncertainty remaining in Z_1 once we have observed the previous sales X_1, X_2. This might appear to be good news: we are, after all, hoping to reduce our uncertainty about future sales by linear fitting on these two explanatory quantities. However, let us look a little more closely at the implications. Z_1 is proportional (except for a constant) to total sales: $Y_1 + Y_2 = 8.944Z_1 + 200$, so that one implication of our prior specification is that we shall have no uncertainty about $Y_1 + Y_2$ after we have observed X_1 and X_2. Indeed, as the adjusted expectations of Y_1, Y_2 are given above as $\mathrm{E}_d(Y_1) = 108.75$ and $\mathrm{E}_d(Y_2) = 90.25$ respectively, we shall apparently know certainly that $Y_1 + Y_2$ will be $108.75 + 90.25 = 199$. Did we really intend our prior specifications to contain the algebraic implication that we will 'know' total future sales in advance? Most likely we did not; and indeed later we actually observe total sales of $y_1 + y_2 = 112 + 95.5 = 207.5$, which flatly contradicts the prior specification, and which resulted in the failure of the consistency check in the previous section.

Now, what has led to this position? To find out, we obtain the adjusted expectations for the canonical quantities Z_1 and Z_2. For simplicity we introduce an obvious notation for the main sums and differences:

$$X^+ = X_1 + X_2, \quad X^- = X_1 - X_2, \quad Y^+ = Y_1 + Y_2, \quad Y^- = Y_1 - Y_2.$$

The adjusted expectations for the canonical quantities are:

$$\mathrm{E}_D(Z_1) = 0.224X^+ - 44.722, \qquad (1.14)$$

$$\mathrm{E}_D(Z_2) = 0.056X^-.$$

The resolution $\mathrm{R}_D(Z_1) = 1$ corresponds to having an adjusted variance of zero for $\mathrm{E}_D(Z_1)$, shown as (1.14), so that the correlation between Z_1 (where $Z_1 \propto Y^+$)

and $\mathrm{E}_D(Z_1)$ (where $\mathrm{E}_D(Z_1) \propto X^+$) must be equal to one. Thus, X^+ and Y^+ have a prior correlation of one, and this explains why Y^+ becomes 'known' as soon as we observe x^+.

Now, while this was a logical consequence of our prior specification, it is quite possible that we had not realized, when we made our pairwise prior correlation specifications, that we were building such a strong degree of dependency between X^+ and Y^+. Indeed, it will usually be the case, particularly when we come to specify beliefs over large, complex and highly interdependent collections of quantities, that our initial prior specifications will have surprising and counter-intuitive consequences, which may cause us to reconsider the basis for our specifications. It is for this reason that it is vital to carry out a global analysis, by generating and examining the canonical structure, to ensure coherence and consistency over and between belief specifications and data. In particular, many defects are not discovered if we carry out analyses piecemeal – for example, nearly all of the analyses carried out in §1.4.3 to §1.4.10 are unremarkable when Y_1 and Y_2 are considered separately, but are revealed to be dubious when we analyse them as a collection. We did receive a hint of the underlying problem earlier, in §1.4.5, where we noticed the singularity in the adjusted variance matrix. Singularities showing up here are directly related to finding canonical resolutions equal to one.

In this particular example, the canonical quantities Z_1, Z_2 are the suitably centred and scaled versions of Y^+ and Y^-. Because of the symmetries involved in the prior specification, the **canonical data quantities** $\mathrm{E}_D(Z_1), \mathrm{E}_D(Z_2)$ are likewise the suitably centred and scaled versions of X^+ and X^-. Note that these are also uncorrelated. In later chapters we shall discuss in detail the use of such canonical structures and explain the relationship with classical canonical correlation analysis.

1.4.13 Modifying the original specifications

In this case, let us suppose that we reconsider our prior specifications. There are many changes that we might make. Suppose, for simplicity, that we decide not to change our prior means and variances for the four sales quantities, but just to weaken one or two of the correlations. In terms of the four sums and differences, the original prior correlation matrix was:

$$\begin{array}{lllll}
 & X^- & X^+ & Y^- & Y^+ \\
X^- & 1 & & & \\
X^+ & 0 & 1 & & \\
Y^- & 0.5 & 0 & 1 & \\
Y^+ & 0 & 1 & 0 & 1
\end{array}$$

Inspecting the matrix, suppose we decide that it is appropriate to weaken the correlation between X^+ and Y^+ to 0.8. With this change, the prior correlation matrix over sales becomes

	X_1	X_2	Y_1	Y_2
X_1	1			
X_2	−0.60	1		
Y_1	0.56	−0.24	1	
Y_2	−0.24	0.56	−0.60	1

so that the actual effect is to decrease generally all the correlations between the sales quantities.

1.4.14 Repeating the analysis for the revised model

We now repeat our analysis with the modified belief specifications. The results are rather similar, and have similar interpretations. The adjusted expectations are now

$$\begin{aligned} \mathrm{E}_D(Y_1) &= 100 + 1.3(X_1 - 100) + 0.3(X_2 - 100), \qquad (1.15)\\ \mathrm{E}_D(Y_2) &= 100 + 0.3(X_1 - 100) + 1.3(X_2 - 100), \end{aligned}$$

so that $x_1 = 109, x_2 = 90.5$ yields observed adjusted expectations of

$$\mathrm{E}_d(Y_1) = 108.85 \quad \text{and} \quad \mathrm{E}_d(Y_2) = 90.35.$$

These are about the same as for the original prior specifications. As before, the adjusted variances are the same for the two products,

$$\mathrm{Var}_D(Y_1) = \mathrm{Var}_D(Y_2) = 67.2,$$

so that the variance resolutions are 32.8%. Compared to the original specifications, the weakening of the underlying correlations leads to the explanatory quantities being less informative for future sales. The standardized changes in expectation (prior to adjusted) are $\mathrm{S}(\mathrm{E}_d(Y_1)) = 1.55$ and $\mathrm{S}(\mathrm{E}_d(Y_2)) = -1.69$, a little larger than before. Finally, when we observe $y_1 = 112$ and $y_2 = 95.5$, the standardized changes from adjusted expectation to observed are 0.38 and −0.63 respectively. Summaries of the basic adjustments are shown in Table 1.2.

In terms of the vector B, the decomposition of the prior variance matrix into unresolved and resolved portions is now (with the correlation matrices shown underneath),

$$\begin{aligned} &\text{Variances:} \quad \begin{bmatrix} 100 & -60 \\ -60 & 100 \end{bmatrix} = \begin{bmatrix} 67.2 & -52.8 \\ -52.8 & 67.2 \end{bmatrix} + \begin{bmatrix} 32.8 & -7.2 \\ -7.2 & 32.8 \end{bmatrix}, \\ &\text{Correlations:} \quad \begin{bmatrix} 1 & -0.6 \\ -0.6 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & -0.79 \\ -0.79 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & -0.22 \\ -0.22 & 1 \end{bmatrix}, \end{aligned}$$

so that unlike for the first prior specification, there has been some alteration to the covariance structure for the residual portions of Y_1 and Y_2. The understanding of such changes to the covariance structure is a matter we defer until later.

Table 1.2 Adjusting future sales Y_1, Y_2 by previous sales: summary for the modified structure, giving expectations E(·), variances Var(·), and standardized changes S(·).

	Initial	=	Adjusted expectation	+	Adjusted version
	Y_1	=	$0.3X_1 + 1.3X_2 - 100$	+	$Y_1 - (0.3X_1 + 1.3X_2 - 100)$
Prior E(·)	100	=	100	+	0
Prior Var(·)	100	=	32.8	+	67.2
Data	112	=	108.85	+	3.15
Change S(·)	1.2		1.55		0.38
	Y_2	=	$0.3X_1 + 1.3X_2 - 100$	+	$Y_2 - (0.3X_1 + 1.3X_2 - 100)$
Prior E(·)	100	=	100	+	0
Prior Var(·)	100	=	32.8	+	67.2
Data	95.5	=	90.35	+	5.15
Change S(·)	−0.45		−1.69		−0.63

For the modified model, we recalculate the canonical structure. The two canonical directions are as in (1.12) and (1.13), with corresponding canonical resolutions $\mathrm{R}_D(Z_1) = 0.64$ and $\mathrm{R}_D(Z_2) = 0.25$. It follows that we expect to 'explain' 64% of the variation in the direction/linear combination $Z_1 \propto Y^+$, and this is the most we can learn about any linear combination of the two future sales quantities. Otherwise, the canonical structure is as before.

The canonical structure helps us to understand the implications of our belief specifications. There are two ideas. The first is that we examine the implications of our belief specifications as they affect variance reduction, and the second is that we do this globally, i.e. simultaneously over all linear combinations of interest, thereby taking account of the relationships expressed between the quantities being predicted. Our unknowns have been reorganized as a canonical structure which has two directions, scaled so that the prior variance in each is one, and so that the removed variance in each is the corresponding canonical resolution. Consequently we will talk of the global structure as having initial uncertainty $1 + 1 = 2$ and resolved uncertainty $0.64 + 0.25 = 0.89$, with resolution averaged over the structure evaluated as $0.89/2 = 0.445$. This single number, which we call the **system resolution** for our collection B of future sales quantities, is a simple quantification of the value of the information for the entire collection B. We treat the system resolution just as we treat resolutions for individual quantities such as Y_1. That is, a system resolution of zero implies that the information contains no potential to reduce uncertainties in the collection by linear fitting, whereas a system resolution of one implies that the information precisely identifies all the elements of the collection B. In this way we begin to distance ourselves from the idea that the individual quantities are the fundamentals of interest, and approach instead the idea that the **collections** constitute the fundamentals of interest. This blurring of the distinction between single quantities and collections of them has many advantages, particularly as the dimensionality of a problem increases.

1.4.15 Global analysis of collections of observations

In previous sections we saw that piecemeal analyses for individual quantities such as Y_1 provided little or no evidence of the serious flaws present in the prior belief specification; these flaws were revealed only by calculating and interpreting the underlying canonical structure. In a Bayes linear analysis we assess both the expected value of information sources and diagnostics (such as standardized changes) comparing expected to actual behaviour. Therefore, the question arises: is it sufficient to examine standardized changes for the single elements of a collection, or, analogous to the underlying canonical structure, is there a more informative underlying diagnostic structure? Recall that one motivation for calculating the canonical structure was to find the linear combination with maximum variance reduction. Suppose, analogously, that we calculate the linear combination Y^* with the largest squared change in expectation, relative to prior variance. For the observations $x_1 = 109$, $x_2 = 90.5$, this turns out to be

$$Y^* = 0.0478Y_1 - 0.0678Y_2 + 2.0000.$$

This linear combination, which has been centred so that it has prior expectation zero, has adjusted expectation $\mathrm{E}_D(Y^*) = 1.078$ and, for a reason we shall come to, a prior variance also of 1.078. Thus, the largest change in expectation from prior to adjusted for any linear combination of the future sales quantities is about $\sqrt{1.078} = 1.038$ prior standard deviations. It appears that the interplay between prior specifications and the data used to compute adjusted expectations is about as expected. As Y^* has been deliberately chosen to maximize the squared standardized change in expectation, we now describe how to assess the magnitude of the maximal change associated with it.

It turns out that Y^* has a unique and important role to play in Bayes linear analysis, and so we introduce a notation and a name for it. For a collection B being adjusted by a further collection D observed to be d, we call the linear combination in B with the largest standardized squared change in expectation the **bearing**, and we use the notation $\mathbb{Z}_d(B)$ for it. It is a simple linear combination of the quantities being predicted (here, Y_1 and Y_2), with the coefficients being functions of the data used to generate the observed adjusted expectation (here, x_1 and x_2). The bearing has two useful properties.

1.4.15.1 Summary of direction and magnitude of changes

The bearing summarizes the direction and magnitude of changes between prior and adjusted beliefs in the following sense: for any quantity Y constructed from the elements of the collection B, the change in expectation from prior to adjusted is equal to the prior covariance between Y and the bearing $\mathbb{Z}_d(B)$ so that $\mathrm{E}_d(Y) - \mathrm{E}(Y) = \mathrm{Cov}(Y, \mathbb{Z}_d(B))$. In our example it is simple to illustrate this result: we have

$$\mathbb{Z}_d(B) = 0.0478Y_1 - 0.0678Y_2 + 2,$$

so that

$$\begin{aligned}\mathrm{Cov}(Y_1, \mathbb{Z}_d(B)) &= \mathrm{Cov}(Y_1, 0.0478Y_1 - 0.0678Y_2 + 2)\\ &= 8.85 = 108.85 - 100\end{aligned}$$

and

$$\mathrm{Cov}(Y_2, \mathbb{Z}_d(B)) = -9.65 = 90.35 - 100.$$

Changes in expectation for other linear combinations, such as Y^+ and Y^-, are obtained as easily. For example,

$$\begin{aligned}\mathrm{E}_d(Y^+) &= \mathrm{Cov}(Y^+, \mathbb{Z}_d(B)) = -0.8,\\ \mathrm{E}_d(Y^-) &= \mathrm{Cov}(Y^-, \mathbb{Z}_d(B)) = 18.5.\end{aligned}$$

In particular, recalling that we noticed above that Y^* has a prior variance equal to its change in expectation, 1.078, we now observe that this is explained because

$$\mathrm{E}_d(\mathbb{Z}_d(B)) - \mathrm{E}(\mathbb{Z}_d(B)) = \mathrm{Cov}(\mathbb{Z}_d(B), \mathbb{Z}_d(B)) = \mathrm{Var}(\mathbb{Z}_d(B)).$$

1.4.15.2 Global diagnostic

The bearing provides a global diagnostic which gives a guide as to how well the data agree with the prior information. We have already seen that $\mathbb{Z}_d(B)$ is the linear combination having the largest squared change in expectation, relative to prior variance. We will call this change, which we have seen is just $\mathrm{Var}(\mathbb{Z}_d(B))$, the **size of the adjustment**, and introduce the notation $\mathrm{Size}_d(B)$ for it. It is natural to compare this maximum data effect with our expectation $\mathrm{E}(\mathrm{Size}_D(B))$ for it, where expectation is with respect to the data quantities and prior to them being observed. This expectation turns out to be

$$\mathrm{E}(\mathrm{Size}_D(B)) = \mathrm{E}(\mathrm{Var}(\mathbb{Z}_D(B))) = \sum_i \mathrm{R}_D(Z_i),$$

i.e. the sum of the canonical resolutions. In our example, the size of the adjustment and its prior expectation are

$$\mathrm{Var}(\mathbb{Z}_d(B)) = 1.078,$$
$$\mathrm{E}(\mathrm{Var}(\mathbb{Z}_D(B))) = 0.64 + 0.25 = 0.89.$$

For a simple global diagnostic we calculate $\mathrm{Sr}_d(B)$, the ratio of these quantities, which we call the **size ratio for the adjustment** of B given the observations $D = d$. In our example we obtain $\mathrm{Sr}_d(B) = 1.078/0.89 = 1.21$. This ratio has expectation one. Large size ratios indicate larger than expected changes in expectation, suggesting that the data are in sharp disagreement with our prior specifications. Small size ratios indicate smaller changes in expectation than expected and may imply that our prior variance specifications were too large. In our example, the size ratio is fairly close to one, suggesting little conflict between our prior information and the observations.

1.4.16 Partial adjustments

We have so far addressed the adjustments of both single quantities and collections by a single collection of information sources. We now move on to explore the partial effects and implications of individual pieces of information. In the following example, each ‘information source’ will be a single random quantity, but the approach works in just the same way when the individual information sources are themselves collections of quantities. Some of the reasons for studying partial adjustments are as follows. First, at the **design** stage, some of the information sources may be expensive to observe and so there may be advantages in excluding them as predictors if they are not individually valuable in helping to reduce variation in the unknowns. Secondly, at the **analysis** stage, it is valuable to know which aspects of the data have led us to our conclusions. Thirdly, at the **diagnostic** stage, adjustments are usually based on data from different sources which may or may not be in general agreement – for example, the data from one information source may suggest that an adjusted expectation should rise, whilst data from a different information source may suggest the reverse. In such cases it can easily happen that an overall adjustment appears quite plausible, but conceals surprising conflicts between different pieces of evidence. Bayes linear analysis permits us to explore the interactions between the various sources of beliefs and data in a way which highlights any such discordant features.

Key to understanding (linear) partial effects is the notion that one information source is often at least partly a surrogate for another information source. For example, if two vectors U and V are perfectly correlated in the sense that every linear combination constructed from the elements of U is perfectly correlated with some linear combination constructed from the elements of V, then we could essentially throw away V as U carries all the relevant information. Thus, when U and V are correlated, there will be a portion of V which is irrelevant when we also have U, and vice versa. We introduced in §1.4.4 the notion of, and notation for, the decomposition of a single random quantity into an adjusted expectation plus an adjusted version. We now extend this notation to vectors of random quantities. That is, we write

$$U = \mathrm{E}_V(U) + [U - \mathrm{E}_V(U)] = \mathrm{E}_V(U) + \mathbb{A}_V(U).$$

Informally, in a linear framework, $\mathrm{E}_V(U)$ and $\mathbb{A}_V(U)$ are respectively (1) the portion of the information source U that is also carried by V, and (2) the residual portion of U not duplicated by any part of the information source V.

Before we illustrate the Bayes linear approach to design via partial adjustment, it may be helpful to consider the usefulness of summaries of partial effects in the traditional context of stepwise linear regression. In stepwise regression the usual setting is that of one or more response variables with a large number of explanatory variables, where it is desired to determine a small subset of explanatory variables according to some criterion – such as the explanation of a given percentage of variation in the response variables. Two simple approaches to finding such a subset are forward selection and backward elimination. The former proceeds by beginning with an empty set of explanatory variables and then sequentially adding to this set the explanatory variables which are most helpful in explaining remaining variation in the response variables. The latter proceeds by taking the full set of explanatory variables and then sequentially removing those explanatory variables which are least helpful in explaining variation in the response variables. Both these notions have their analogues in Bayes linear methodology. With regard to forward addition of variables, the partial effect of interest is the extra percentage of variance explained in the response variables. With regard to backward deletion of variables, the partial effect of interest is the reduction in the explained variance of the response variables attributable to removing an explanatory variable.

In our example so far we have used our information sources X_1 and X_2 jointly as D to learn about future sales. Suppose now that we consider how important each is individually in predicting future sales. We adjust first by X_1 and then perform the **partial adjustment** by $\mathbb{A}_{X_1}(X_2)$, the adjusted version of X_2 given X_1, which is the portion of X_2 which has not already been contributed to the adjustment by X_1.

Details of the resulting variance resolutions are shown in Table 1.3. For example, when X_1 alone is used, the expected variance resolution in Y_1 is $\mathrm{R}_{X_1}(Y_1) = 0.3136$, rising to $\mathrm{R}_D(Y_1) = 0.3280$ when X_2 is also used. The **partial resolution** contributed by $\mathbb{A}_{X_1}(X_2)$ is thus, by subtraction,

$$\mathrm{R}_{\mathbb{A}_{X_1}(X_2)}(Y_1) = 0.0144.$$

Table 1.3 shows clearly that X_1 is mainly informative for Y_1, and that the residual portion of X_2 having taken into account X_1 has little extra explanatory power. For

Table 1.3 Variance implications for individual quantities and their collection.

	Resolution given X_1 $\mathrm{R}_{X_1}(\cdot)$	Resolution given X_1 and X_2 $\mathrm{R}_{X_1 \cup X_2}(\cdot)$	Partial resolution $\mathrm{R}_{\mathbb{A}_{X_1}(X_2)}(\cdot)$
Y_1	0.3136	0.3280	0.0144
Y_2	0.0576	0.3280	0.2704
B	0.1640	0.4450	0.2810

explaining variation in Y_2, the role is reversed. In the context of this example, if we were particularly interested in predicting sales of Y_1 rather than Y_2, and if X_2 was expensive to measure, we might decide at this stage not to bother observing X_2 but to depend on only observing X_1. Actual design decisions will depend on context and will take into account issues such as the expense of observing quantities such as X_1 and the utility of reducing variation in quantities such as Y_1. If we are concerned with explaining variation globally across the collection B, we notice that the variance resolutions are $\mathrm{R}_{X_1}(B) = 0.1640$ and $\mathrm{R}_{\mathbb{A}_{X_1}(X_2)}(B) = 0.2810$ respectively, indicating that both information sources are valuable.

Given data X_1 alone, the adjusted expectations are

$$\mathrm{E}_{X_1}(Y_1) = 1.12(X_1 - 100) + 100,$$

$$\mathrm{E}_{X_1}(Y_2) = -0.48(X_1 - 100) + 100.$$

Consequently, if we observe X_1 to be larger than expected, the expectations for Y_1 and Y_2 are revised upwards and downwards, respectively. These movements are due to the prior correlations shown in §1.4.13 in that X_1 is positively correlated with Y_1 and negatively correlated with Y_2. The actual observation $x_1 = 109$ gives adjusted expectations of $\mathrm{E}_{x_1}(Y_1) = 110.08$ and $\mathrm{E}_{x_1}(Y_2) = 95.68$. These are standardized changes of ±1.8 standard deviations relative to the variances resolved.

If we now make the partial adjustment by X_2, or rather by the adjusted version $\mathbb{A}_{X_1}(X_2)$, we obtain **partial adjusted expectations** which provide the formulae to update the expectations from the current adjusted expectation (given only X_1) to that based on both X_1 and X_2. In doing so, it is helpful to introduce some extra notation. Let

$$\mathrm{E}_{[X_2/X_1]}(B) = \mathrm{E}_{X_1 \cup X_2}(B) - \mathrm{E}_{X_1}(B)$$

be the partial adjustment of B by X_2 given that we have already adjusted by X_1. Such partial adjustments necessarily have expectation zero. We find that

$$\mathrm{E}_{[X_2/X_1]}(Y_1) = 0.18(X_1 - 100) + 0.30(X_2 - 100),$$

$$\mathrm{E}_{[X_2/X_1]}(Y_2) = 0.78(X_1 - 100) + 1.30(X_2 - 100).$$

In this case, if we observe X_2 to be larger than expected, the partial change in expectation for both Y_1 and Y_2 is upward. As we did observe $x_2 = 90.5$, the partial change in expectation is $0.18(109 - 100) + 0.30(90.5 - 100) = -1.23$ for Y_1

Table 1.4 Exploring the implications of partial adjustment for Y_1 and Y_2.

	Results for Y_1		
	Prior	Given X_1	Given X_1 and X_2
Expectation	100.0	110.08	108.80
Variance	100.0	68.64	67.20
Total variance resolved		31.36	32.80
Change in expectation		10.08	−1.28
Change in variance resolved		31.36	1.44
Squared standardized change in expectation		3.24	1.05

	Results for Y_2		
	Prior	Given X_1	Given X_1 and X_2
Expectation	100.0	95.68	90.35
Variance	100.0	94.24	67.20
Total variance resolved		5.76	32.80
Change in expectation		−4.32	−5.33
Change in variance resolved		5.76	27.04
Squared standardized change in expectation		3.24	1.05

and $0.78(109 - 100) + 1.30(90.5 - 100) = -5.33$ for Y_2. These are standardized changes of 1.03 standard deviations relative to the respective resolutions in variance. A summary for the adjustments is given in Table 1.4. Overall, we notice that the expectation for Y_1 rose and then fell back slightly whilst the expectation for Y_2 fell and then fell again. None of the standardized changes are particularly large and we conclude that the magnitudes of the changes in expectation are in apparent agreement with the prior specification.

Because the initial data source X_1 is uncorrelated with the partial data source $\mathbb{A}_{X_1}(X_2)$, notice how the overall adjusted expectations for Y_1 and Y_2 given in (1.15) have been decomposed into additive initial and partial adjustments. That is, we have

$$\mathrm{E}_D(\cdot) = \mathrm{E}_{X_1}(\cdot) + \mathrm{E}_{[X_2/X_1]}(\cdot).$$

1.4.17 Partial diagnostics

We saw in Table 1.4 that the expectation for Y_1 rose and then fell slightly, so that the two information sources, X_1 and $\mathbb{A}_{X_1}(X_2)$, might be said to have contradictory implications for Y_1, whereas the two information sources are apparently complementary as far as Y_2 is concerned. Obviously we can make similar judgements for whichever quantities are of interest, such as total future sales Y^+, but it is simpler to calculate a global summary of the implication of two sources of information. Recall that in §1.4.15 we introduced the **bearing for the adjustment** to summarize the magnitude and direction of changes in expectation implied by a data source. For a partial adjustment we calculate the **bearing for the partial adjustment**, which summarizes the magnitude and direction of changes in expectation implied by the additional partial information source.

In our example, the initial bearing given data x_1, the partial bearing given extra data $\mathbb{A}_{x_1}(x_2)$, and the overall bearing given all the data $d = x_1 \cup x_2$, are

$$\begin{aligned}
\text{Initial:} \qquad \mathbb{Z}_{x_1}(B) &= 0.1170(Y_1 - 100) + 0.0270(Y_2 - 100) \\
\text{Partial:} \quad \mathbb{Z}_{[X_2/X_1]}(B) &= -0.0692(Y_1 - 100) - 0.0948(Y_2 - 100) \\
\text{Overall:} \qquad \mathbb{Z}_d(B) &= \mathbb{Z}_{x_1}(B) + \mathbb{Z}_{[X_2/X_1]}(B) \\
&= 0.0478(Y_1 - 100) - 0.0678(Y_2 - 100).
\end{aligned}$$

As in §1.4.15, each bearing is associated with a **size ratio** measuring the discrepancy between data and belief specifications taken as a whole across the collection being adjusted. In this example, the size ratios for the initial, partial, and overall adjustments are respectively 3.24, 1.05, and 1.21. None of these, each of which has prior expectation unity, appears particularly large or disturbing, and we might conclude that the changes in expectation implied by the data are in general agreement with the prior specifications.

As a change in expectation for any quantity such as Y_1 can be represented as a covariance between that quantity and a bearing, we also note that the implications of the two data sources for changes in expectation are opposite: typically positive for the first, and typically negative for the second. To formalize this idea, the most useful single summary is the correlation between the bearings for the two data sources, which we call a **path correlation**. In this example, it is

$$\text{PC}(x_1, \mathbb{A}_{x_1}(x_2)) = \text{Corr}(\mathbb{Z}_{x_1}(B), \mathbb{Z}_{[X_2/X_1]}(B)) = -0.3633.$$

The interpretation is that there is a very mild form of conflict between the two information sources.

We have already seen that the standardized changes in expectation at each stage for the two quantities are not too surprising in relation to the variance resolved at each stage. However, we should be aware that an overall adjustment by all the data can mask (either by cancelling out or by averaging) two surprising and/or contradictory changes in belief. As an illustration, we repeat the diagnostic analysis using the canonical structure for the data quantities, which we saw at the foot of §1.4.12 to be the current sales total and sales difference, X^+ and X^-. Thus, we reorganize the data sources to be these canonical data quantities, and use them to make predictions about future sales.

The analysis proceeds as described in previous sections, but we shall not detail it as our interest here is only in the diagnostic evidence. Suppose that we carry out an initial adjustment of B by X^+, and then a further partial adjustment by X^-, which is uncorrelated with X^+, so that we have $\mathbb{A}_{X^+}(X^-) = X^-$. We find that the bearings are

$$\begin{aligned} \text{Initial:} \quad & \mathbb{Z}_{x^+}(B) = -0.01(Y^+ - 200) \\ \text{Partial:} \quad & \mathbb{Z}_{x^-}(B) = 0.0578Y^- \\ \text{Overall:} \quad & \mathbb{Z}_d(B) = 0.0478(Y_1 - 100) - 0.0678(Y_2 - 100), \end{aligned}$$

so that there is a natural and straightforward correspondence between data sources and what the data source is informative for: previous total sales are informative for future total sales, and previous sales differences for future sales differences. Because of the uncorrelatedness of these quantities, observe for example that previous sales totals X^+ are valueless for making linear predictions about a future sales difference, Y^-. The overall bearing $\mathbb{Z}_d(B)$, which is of course the same however we reorganize the information sources, has a corresponding size ratio of 1.21. However, the size ratios for the initial and partial adjustments are respectively 0.0125 and 4.2781. The interpretation here is that the changes in expectation induced by the first data source, X^+, were surprisingly small compared to the expected level of variance explained, whereas the changes in expectation induced by the second data source, X^-, were perhaps disturbingly large. A plausible explanation would be that we overstated our prior variability for the sales totals, and that we understated variability for the sales differences, or perhaps that there are errors in the data. In such cases, we might choose to re-examine our prior specifications and the data. Note that, as will often be the case, diagnostic inspection based on the canonical structure gives a clearer picture of potential problems with the overall prior formulation than is obtained by inspection of the adjustments of the original quantities.

1.4.18 Summary

A good analysis of even simple problems such as these requires the knowledgeable use of effective tools. Our analysis here is incomplete as we have only introduced some of the basic machinery of the Bayes linear approach, and yet we have shown how fairly simple ideas and procedures lead directly into the heart of a problem, offering tools that work as well for collections as they do for single quantities, and that reveal quickly the important aspects of a combined belief and data structure. We could possibly have made a more detailed prior specification. However, by concentrating on the reduced belief specifications required for the second-order structure we have been able to apply a simple and efficient methodology under which we can control input requirements, and within which the implications of the belief specifications and any observations can be readily discerned. Various aspects of the Bayes linear analysis are thus revealed: straightforward specification of genuine beliefs, exploration of their implications, their adjustment using data, and diagnostics comparing expected to actual behaviour. This methodology works in essentially the same way as we increase the number of quantities in the problem, in which case we will find that the role of the canonical structure becomes increasingly important in clarifying the effects of complex belief adjustments.

1.5 Overview

The Bayes linear approach has been developed to the level where it is usable as a general framework within which to develop statistical methodology. As with any such methodology, much work may be required to bring the approach to bear on particularly challenging practical problems. However, the basic elements of the approach are sufficiently well developed to merit a unified exposition. Our intention, in this book, is to present in a systematic way those central methodological features that we consider to be both essential for and distinctive to the Bayes linear approach. Thus, we do not address the many aspects of belief specification, statistical modelling and data analysis which are common to our approach and other views of statistical analysis. Nor do we attempt to summarize all of the ways in which moment specification and analysis are currently exploited within statistical methodology. Instead, by concentrating on the essentials of the approach, we aim to give at least the outline of a unified methodology for belief analysis from a particular subjectivist viewpoint based on partial belief specification taking expectation as primitive. Whether we consider this approach as (the skeleton of) a complete methodology of itself or as part of a much larger toolkit of approaches to belief modelling and analysis will depend both on our philosophical viewpoint and on the types of problem which we wish to address.

The organization of this book is as follows. In Chapter 2, we introduce the ingredients which we will blend in later chapters, namely prior means, variances and covariances, assessed as primitive quantities. We give a brief introduction to the idea of expectation as primitive, and discuss, by example, some simple approaches to prior specification for means, variances and covariances.

The basics of our approach are threefold: (i) we specify collections of beliefs and analyse how we expect beliefs to change given our planned data collection; (ii) we collect information and analyse how our beliefs have actually changed; (iii) we compare, diagnostically, expected to actual changes in our beliefs. Step (i) is addressed in Chapter 3, where we explain the basic operations within our approach, namely the adjustment of collections of expectations and variances, by linear fitting on data. We develop the basic properties of belief adjustment and describe the natural geometric setting for the analysis. A general construction is introduced, namely the belief transform, for interpreting collections of belief adjustments through the eigenstructure of the transform.

We address steps (ii) and (iii) of our general approach in Chapter 4, which is concerned with interpretation and diagnostic evaluation of the observed belief adjustment given data. In particular, we describe the construction and interpretation of the bearing for a belief adjustment, which is a form of linear likelihood for the analysis, which summarizes the overall direction and magnitude of a collection of adjustments.

Usually, our information comes from different sources: for example, there may be different time points, different populations, different types of quantity. It is useful to identify how much information we expect from each source, and then to consider whether the various data sources are giving consistent or a contradictory information. In Chapter 5, we apply the three-step programme – (i) interpret expected adjustments, (ii) interpret actual adjustments, (iii) compare actual to interpreted effects – when the data have been divided into portions. We therefore consider partial belief adjustments and develop the corresponding partial belief transforms and partial bearings for an adjustment carried out in stages.

Exchangeability (the property that beliefs over a collection of objects would not be affected by permutation of the order of the objects) is a fundamental subjective judgement underlying many statistical applications. In principle, exchangeability judgements allow us to carry out statistical modelling purely in terms of our judgements over observables. Unfortunately, in the usual Bayes formalism, this is very difficult, and exchangeability tends to be hidden from view. Because of our simplified approach to belief specification, however, it is both feasible and natural to build statistical models directly from second-order exchangeability judgements over observables. This process is covered in Chapter 6, where we develop and interpret the representation theorem for second-order exchangeable random quantities. Chapter 6 is also concerned with how to adjust beliefs over the resulting exchangeability models. We derive useful general results which greatly simplify the analysis of such models, through the special properties of the corresponding belief transforms. In Chapter 7, we extend such analyses to cover collections of data which are individually second-order exchangeable, and which satisfy natural second-order exchangeability relationships between each pair of collections. In Chapter 8, we address the issues that arise in learning about population variances from exchangeable samples.

To this point, we have treated a particular type of belief transform as our basic interpretative tool for analysing collections of belief changes. However, this type of transform is itself a special case of a much wider class of transforms, which are examined in Chapter 9, all of which are based on comparisons between collections of variance and covariance specifications. We give the general construction for such transforms, and illustrate the approach with various problems of comparison over models and designs.

Graphical models are a powerful tool for graphically representing and evaluating our beliefs. Bayes linear graphical models, covered in Chapter 10, perform this task for describing and manipulating our second-order specifications. We may also display quantitative information, expressing our three-step sequence – expected effects, observed effects and their comparison – in a natural way on the diagram. Thus, the diagrams express both the modelling and the analysis of beliefs. Further, the local computation properties of these models allow us to tackle large problems in a straightforward and systematic way.

In Chapter 12, we cover the technical material that we need for efficient implementation of the Bayes linear approach, assuming a somewhat higher level of knowledge of matrix algebra than in the rest of the book. The matrix algebra required is covered in Chapter 11.