Bayes Linear Methods II
An example with an introduction to [B/D]

David Wooff
University of Durham

[B/D] Home Page: http://fourier.dur.ac.uk:8000/stats/bd/

Abstract

This article is one of a series of papers and technical reports describing Bayes linear methods. Our purpose is to introduce both the approach and the computer implementation of the approach, the programming language [B/D], by way of a simple (but genuine) example. The three parts of this article contain (i) the establishment of the example, concerning the oral glucose tolerance test used in diagnosing diabetes; (ii) a Bayes linear analysis of the example to explore some of the main features of the approach; and (iii) a tutorial in the use of the subjectivist statistical programming language [B/D] used to perform the analysis for the example.

Part I
Introduction and motivation

1 Organisation of this document

Several related documents explore (or will explore) in detail various aspects of Bayes linear methods. In particular, this document accompanies [4], which introduces the basic machinery of the approach. We aim to consider in further documents the implementation of the theory from a computational viewpoint; the elicitation of beliefs; and other aspects of Bayes linear methods. An earlier incarnation of the [B/D] package was considered in [2] and [6]; and a resumé of the approach and the package is given in [7]. More recently, [8] is a technical reference manual, and [5] considers in some detail advanced [B/D] macros used to help analyse an industrial problem.

Here we are concerned with a gentle introduction to some of the features of the approach and to the package used to perform the analyses. The theoretical machinery of the approach is considered in detail in the companion document [4], and this document is intended to provide a tutorial on the interpretation and computation of the material contained therein. For our main illustration we take a genuine problem concerning the diagnoses of diabetes in the elderly. We will assume that we have already undergone the effort necessary for the consideration of all the inputs necessary for our illustration, so that the belief specifications are ready and waiting to be used. This is not to understate the importance of such belief elicitation aspects, but because we concentrate on those features of our approach elsewhere. Similarly, we will not restate the theory and interpretation found in the companion document [4], or various implementational aspects.

This document consists of three parts. In the second part we take a simple (but genuine) problem and show how we apply Bayes linear methods to perform adjustments of belief (informally, the revision process) and to assess the implications of both the belief specifications and the actual observations for the problem. In this way, we explore the central features of the Bayes linear approach.

In part 3 we show how our computer program [B/D] may be used to perform all the analyses considered in part 2, and in so doing we provide a tutored introduction to the use of the [B/D] programming language. In this first part we motivate and introduce the example that we use to illustrate both the methodology and the programming language. It is our intention that the part containing our illustration of the methodology via example stands separate from the final, programming part in case you have no access to (or interest in!) the [B/D] program.

2 Why Bayes linear methods?

Our approach is characterised by several features:

- It is subjectivist: we deal with uncertain quantities (like the average daily temperature during next year's summer); we hold beliefs about such quantities (we expect it to be cold again); and we can measure related quantities (such as whether it snows on Chrismas day). We combine such information in order to help us revise our beliefs about these uncertain quantities.
- It is based upon expectation as a primitive: we deal with expectations of uncertain quantities, rather than probabilities of events. (If this worries you as seeming rather restrictive, recall that probabilities are expectations of indicator functions: let A be any event, and define I_A to be the random quantity which takes the value $I_A = 1$ if A occurs, and $I_A = 0$ otherwise. Then $Pr(A) = \mathrm{E}(I_A)$.)
- It is practicable: we demand of you only that you specify the information (in the form of expectations, variances and covariances) that you are both willing and able to provide. In contrast, a standard Bayesian treatment typically requires the provision of full multidimensional probability distribution over all the quantities of interest (whether or not this can be done meaningfully).
- It is simple; traditional Bayesian methods are frequently extremely demanding in terms of the computational effort needed for an analysis, whereas the Bayes linear approach preserves via its linearity a natural tractability.
- It is insightful: the information contained within our constructions is readily accessible, and some of the tools that we provide yield insights that are difficult to obtain under other approaches. This follows partly because a traditional Bayesian approach, for example, can involve a mass of specifications (much of it arguably arbitrary) and calculations from which it is difficult to extract revealing summaries.
- It subsumes the traditional Bayesian analysis as a special limiting case (in the sense of expectations over indicator functions).
- The Bayes linear approach can be considered to be an approximation to the full Bayesian treatment in the case of certain limited specifications.
- It is logically supportable. (See [3] for a discussion of underlying philosophical issues.)

Thus, in our view our approach is logically founded; tractable and simple to apply; easily understood and well-defined; genuine and achievable.

3 Establishing the example

Our example concerns the **oral glucose tolerance test** (ogt-test), used as an indicator of diabetes. The ogt-test is intended to measure the time taken to absorb a given quantity of sugar, with high values indicating diabetes. The test involves measuring the blood-glucose level in mmol/litre before and after a fixed quantity of sugar is taken orally (after fasting for twelve hours). Typically the blood glucose level is measured immediately prior to, and two hours after, the sugar is swallowed. Additionally, measurements might be made at intervening half-hour points. The most frequently utilised measurements are the fasting and two-hour measurements, taken before and two hours after ingestion of the sugar, and a well-established diagnosis of diabetes depends only upon these two measurements. Threshholds for the various diagnoses are given in figure 1.

We examine the efficacy of the ogt-test in relation to elderly people (for our purposes, people at least 60 years old). Previous trials of the ogt-test dealt largely with younger people, but there is a suspicion that the threshholds that indicate diabetes in a young person are inappropriate for older people. In particular, it is

Figure 1: Oral glucose tolerance tests: diagnosis threshholds

blood-glucose level mmol/litre	diagnosis for fasting measurement	blood-glucose level mmol/litre	diagnosis for 2-hour measurement
under 7 more than 7	Healthy Diabetes	under 7 7 to 10 more than 10	Healthy Impaired glucose tolerance Diabetes

possible that the two-hour diagnosis threshhold is set too high for the elderly because their bodies react more slowly in general, and so take longer to absorb sugar. As a consequence, it is suspected that the ogt-test might often misdiagnose impaired glucose tolerance or diabetes in healthy elderly patients.

For our chosen scenario we suppose that an elderly lady doctor is interested in the adequacy of the test for older people, and that she decides to administer the ogt-test on herself to obtain further information. To help organise her thoughts, she writes down the quantities B_1 and B_2, representing respectively the blood glucose levels measured before and two hours after swallowing the sugar. She thinks of each as a measurement on a randomly chosen elderly person, presumed healthy and non-diabetic. Corresponding to these quantities, she considers the measurements that she will make upon herself, and writes these down as D_1 and D_2. Thus D_1 and D_2 are the (observable) before-and-after measurements that she makes on herself; whereas B_1 and B_2 are the (unobservable) before-and-after measurements for an elderly person picked at random. Our Doctor's aim is to learn about B_1, B_2, and various linear combinations such as $B_2 - B_1$, via D_1 and D_2. Informally you might think of the former pair as characterising a typical member of the (elderly) population; and the latter pair as a "sample" from that population.

At this point we will introduce some notation. We only distinguish between collections of quantities, which we term **bases**, and the more formal organisation of such collections as vectors of quantities, when this is not clear from the context. Thus we will refer both to the bases

$$\begin{aligned} B &= \{B_1, B_2\}, \\ D &= \{D_1, D_2\}. \end{aligned}$$

and to the vectors B and D defined to be

$$B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \quad D = \begin{bmatrix} D_1 \\ D_2 \end{bmatrix}.$$

As we shall see, our approach allows us to obtain general impressions about collections of quantities as well as individual quantities, and when taking such a wider view we use the following notation:

- $\langle B \rangle$ means the *vector space* spanned by the elements of B so that here $\langle B \rangle$ means all the possible linear combinations of B_1 and B_2. For example, we might be particularly interested in the difference $B_2 - B_1$.
- $[B]$ means a **belief structure** generated over $\langle B \rangle$ by a specification of expectations, variances and co-variances for the elements of B.

The minimum specifications that we require for an analysis are expectations, variances and covariances over all the quantities of interest. If we intend (as we do) to produce adjusted expectations (informally, revisions) then we need some observations as well.

3.1 Expectation specifications

Our Doctor specifies 4.16 as her expectation for the fasting blood-glucose level of a healthy elderly person. This value is suggested by previous studies on healthy young people, and the Doctor's belief that the fasting blood-glucose level is about the same for both the young and the elderly.

However, the Doctor's expectation for the two-hour blood-glucose level for a healthy elderly person is 6.25, somewhat larger than the value of about 5.5 suggested by previous studies on the young. The Doctors's expectation specifications are thus as follows.

$$\mathrm{E}(B) = \begin{bmatrix} \mathrm{E}(B_1) \\ \mathrm{E}(B_2) \end{bmatrix} = \begin{bmatrix} 4.16 \\ 6.25 \end{bmatrix}.$$

$$\mathrm{E}(D) = \begin{bmatrix} \mathrm{E}(D_1) \\ \mathrm{E}(D_2) \end{bmatrix} = \begin{bmatrix} 4.16 \\ 6.25 \end{bmatrix}.$$

Notice that the Doctor specifies the same expectations and covariances about B_1 and B_2 as she does about D_1 and D_2. This is because at this stage the doctor believes herself to be quite typical.

3.2 Covariance specifications

After much reflection, the Doctor specifies variances and covariances over all the quantities of interest as follows[1]:

$$\mathrm{Var}(B) = \begin{bmatrix} \mathrm{Var}(B_1) & \mathrm{Cov}(B_1, B_2) \\ \mathrm{Cov}(B_1, B_2) & \mathrm{Var}(B_2) \end{bmatrix} = \begin{bmatrix} 1.12 & 0.72 \\ 0.72 & 2.43 \end{bmatrix}.$$

$$\mathrm{Var}(D) = \begin{bmatrix} \mathrm{Var}(D_1) & \mathrm{Cov}(D_1, D_2) \\ \mathrm{Cov}(D_1, D_2) & \mathrm{Var}(D_2) \end{bmatrix} = \begin{bmatrix} 1.12 & 0.72 \\ 0.72 & 2.43 \end{bmatrix}.$$

$$\mathrm{Cov}(B, D) = \begin{bmatrix} \mathrm{Cov}(B_1, D_1) & \mathrm{Cov}(B_1, D_2) \\ \mathrm{Cov}(B_2, D_1) & \mathrm{Cov}(B_2, D_2) \end{bmatrix} = \begin{bmatrix} 0.62 & 0.30 \\ 0.30 & 0.43 \end{bmatrix}.$$

To facilitate our analysis, suppose that we calculate the corresponding correlations:

$$\mathrm{Corr}(B, B) = \mathrm{Corr}(D, D) = \begin{bmatrix} 1 & 0.436 \\ 0.436 & 1 \end{bmatrix}.$$

$$\mathrm{Corr}(B, D) = \begin{bmatrix} \mathrm{Corr}(B_1, D_1) & \mathrm{Corr}(B_1, D_2) \\ \mathrm{Corr}(B_2, D_1) & \mathrm{Corr}(B_2, D_2) \end{bmatrix} = \begin{bmatrix} 0.554 & 0.182 \\ 0.182 & 0.177 \end{bmatrix}.$$

Some of the principal features of the Doctor's variance and covariance specifications are as follows:

- Assessing the expectation $\mathrm{E}(B_2) = 6.25$ and variance $\mathrm{Var}(B_2) = 2.43$ specified for B_2, we observe that these specifications reflect the Doctor's belief that the ogt-test tends to misdiagnose many elderly patients in that 6.25 is quite close (less than half a standard deviation distant) to the threshhold of 7.0 for a diagnosis of impaired glucose tolerance.
- She is less sure about the two-hour measurements than she is about the fasting measurements; and she has assigned a correlation of about 0.436 between B_1 and B_2 and between D_1 and D_2.
- Her covariances specified between B and D imply that B_1 and D_1 are moderately related (a correlation of about 0.554), whereas B_2 and D_2 are only weakly related, with the correlation between them being about 0.177. She has also specified about the same degree of relationship (a correlation of about 0.182) between D_1 and B_2. Observe that D_1 is more strongly related to B_2 than is D_2[2]

[1] The specifications are the result of a genuine attempt to express plausible beliefs for the problem, but are not the judgements of an expert.

[2] This is because we consider B_2 as representing the sum of a base level, B_1, plus a difference, $B_2 - B_1$; and D_1 is informative for B_2 via B_1.

Informally, the strength of correlation between two quantities determines the degree to which one quantity can be used to help learn about another. For example, if the Doctor had specified zero correlations between D and B, then observing D would tell us nothing (in a linear analysis) about B. Here she has specified a larger correlation between D_1 and B_2 than between D_2 and B_2, suggesting perhaps that the fasting measurement D_1 will be more informative than the two-hour measurement D_2 in learning about both unknowns B_1 and B_2.

3.3 Data specifications

Finally, suppose that at some point the Doctor undergoes the ogt-test herself, and takes actual measurements d_1 and d_2, representing her fasting blood-glucose level and two-hour blood-glucose level. We write these measurements as a vector d, where

$$d = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} 5.4 \\ 9.8 \end{bmatrix}.$$

Notice immediately that the Doctor is classified as having impaired glucose tolerance because of her reading of 9.8. Indeed, she is very close to being classified as diabetic. Let us reiterate that whilst this example is purely illustrative, it is genuine: these are real observations for a "randomly chosen" elderly person with no history or indication of diabetes.

3.4 Remarks about the example

Bayes linear methods are basically intended for large problems: particularly those that are too big to analyse using a fully specified Bayesian approach. We have chosen for our illustration a small example to keep things simple. Indeed, this simple example is extracted from a very much larger and more complicated problem that we mean to present elsewhere. We intend also to present elsewhere explicit consideration of the belief elicitation process which gave rise to the specifications used here.

In practice, our analysis could take into account the way in which the covariance specifications were produced. For example, our Doctor thought about various underlying quantities such as differences between herself and a young patient. At a later stage, it may then prove possible to return to the way in which the beliefs were generated in order to address queries raised during the general analysis. For example, if the analysis suggests the possibility that she specified too tight a variance for D_2, say, then she may be able to return to the way in which she constructed this variance, and then perhaps to discover a flawed specification for some underlying quantity that she used in the construction. However, whilst valuable, this would overcomplicate our elementary example, and so we omit such considerations.

Part II
The Bayes linear approach by example

4 Adjusting beliefs by data

To this point we have completed our minimal specifications for our example; now we proceed to analyse these specifications. We have uncertain quantities B_1 and B_2, with educated guesses for their locations: expectations $\mathrm{E}(B_1)$ and $\mathrm{E}(B_2)$; and for the accuracy of these guesses: their variances $\mathrm{Var}(B_1)$ and $\mathrm{Var}(B_2)$. We have also the observable quantities D_1 and D_2; expectations and variances for them; and covariances linking B_1 and B_2 with D_1 and D_2. Additionally we have some data on D_1 and D_2. The learning process essentially consists of modifying our expectations for B_1 and B_2, and of improving the accuracy of these expectations in the sense of reducing variances, in the light of the information contained in $[D]$. The terms we use for such modified expectations and variances are **adjusted expectations** and **adjusted variances**, and inter alia we obtain them by **adjusting the belief structure** $[B]$ **by the belief structure** $[D]$. Recall that by belief structure we mean the entirety of specifications over a particular base: essentially the covariance matrix and the expectations for the quantities in the base. One belief structure is adjusted by another via covariances specified between the two underlying bases.

Whilst one of the aims of the analysis is to modify expectations and variances for B in the light of data, let us remember that before we see any data, part of *our* learning process is to assess exactly how the data will be used when it comes. To use the analogy of a traditional statistical estimation procedure, we usually wish to examine not only the "estimate" but also the "estimator" and its properties. Following such examination, when the data arrives we obtain estimates and then check for consistency between what we expected to happen, and what actually happened.

4.1 Adjusted expectations

The adjusted expectations are approximately

$$\begin{aligned}
\mathrm{E}_D(B_1) &= 0.59D_1 - 0.05D_2 + 2.04 \\
\mathrm{E}_D(B_2) &= 0.19D_1 + 0.12D_2 + 4.70.
\end{aligned}$$

These quantities, which are constructed solely from belief specifications, are functions of the (unobserved) data; informally they are akin to estimators in the traditional sense.[3] Generally speaking, it is difficult to tell whether a small coefficient truly indicates unimportance because of the different expectations and scalings of D_1 and D_2. For this reason, we also examine the standardised adjusted expectations, being approximately

$$\begin{aligned}
\mathrm{E}_D(B_1) &= 0.62D_1^* - 0.08D_2^* + 4.16 \\
\mathrm{E}_D(B_2) &= 0.20D_1^* + 0.19D_2^* + 6.25,
\end{aligned}$$

$$\text{where } D_i^* = \mathrm{S}(D_i) = \frac{D_i - \mathrm{E}(D_i)}{\sqrt{\mathrm{Var}(D_i)}}.$$

$\mathrm{S}(\cdot)$ is our general notation for the standardised representation of a quantity. Now each coefficient multiplies a quantity that has expectation zero and variance unity, so that the coefficients are more readily comparable. (The constants added in each case are the initial expectations for B_1 and B_2.) We see that the adjusted expectation for B_1 depends essentially on D_1, plus a base value of 4.16; whereas the adjusted expectation for B_2 depends upon a rather larger base value, plus essentially an average of the before-and-after blood-glucose readings.

[3]For example, for a traditional regression of Y on $X_1, \ldots, X_k$, various linear combinations of the X_i's give estimators for Y. Each such estimator possesses properties which we assess in order to determine whether the estimator is likely to yield good estimates, and so forth.

4.2 Adjusted variances

Now, we are interested not only in how we obtain our adjusted expectations for B_1 and B_2, but also in how valuable the adjustment is in terms of reducing uncertainty. To this purpose we examine the adjusted variances: in terms of our general notation we have approximately

$$\begin{aligned}\mathrm{Var}_D(B_1) &= 0.77\\ \mathrm{Var}_D(B_2) &= 2.32.\end{aligned}$$

Here, $\mathrm{Var}_D(X)$ is our notation for the uncertainty remaining in X in the light of the information contained in $[D]$, so we expect uncertainty in B_1 to reduce from $\mathrm{Var}(B_1) = 1.12$ to $\mathrm{Var}_D(B_1) = 0.77$; and the uncertainty in B_2 to reduce from $\mathrm{Var}(B_2) = 2.43$ to $\mathrm{Var}_D(B_2) = 2.32$. We call the difference between these two values the resolved variance, and we also determine the **resolution**, which is a scale-free measure of the reduction in variance. For the belief structure taken as a whole we can determine analagous results.

For each quantity, we can decompose the initial variation into portions remaining and resolved. For B_1 the decomposition is

$$\begin{array}{ccccc}\mathrm{Var}(B_1) & = & \mathrm{Var}_D(B_1) & + & \mathrm{RVar}_D(B_1)\\ 1.12 & = & 0.77 & + & 0.35\\ \text{initial uncertainty} & = & \text{remaining uncertainty} & + & \text{resolved uncertainty}\end{array}$$

For B_2 the decomposition is

$$\begin{array}{ccccc}\mathrm{Var}(B_2) & = & \mathrm{Var}_D(B_2) & + & \mathrm{RVar}_D(B_2)\\ 2.43 & = & 2.32 & + & 0.11\end{array}$$

The resolution for each quantity can lie between zero and unity inclusive, with a resolution of zero implying that we expect the adjustment to be completely uninformative, and a resolution of unity implying that we expect the adjustment to be completely informative. For our two quantities B_1 and B_2 we obtain

$$\begin{aligned}\mathrm{R}_D(B_1) &= 0.3109\\ \mathrm{R}_D(B_2) &= 0.0448\end{aligned}$$

which tells us that by adjusting on $[D]$, we reduce the uncertainty in B_1 by about 31%, and we reduce the uncertainty in B_2 only by about 5%. The adjustment is informative in only a limited sense for B_1, and hardly at all for B_2.

4.3 Adjusting a collection of quantities

We now examine results summarising uncertainties within the entire belief structure $[B]$. By convention we take the initial uncertainty in the belief structure to be the rank of the variance matrix associated with it: the maximal number of mutually uncorrelated quantities in the structure according to the covariances expressed over it[4]. We obtain the adjusted uncertainty for the structure essentially by adding the adjusted variances for these uncorrelated quantities (scaled to have prior variance unity). We take the resolved uncertainty to be the difference between the initial and adjusted uncertainty, and the resolution for the belief structure follows

[4]We identify any quantity having zero variance with zero. Thus, the number of mutually uncorrelated quantities in the structure represents the number of underlying different axes of variation. For example if a belief structure, $[X]$, consisted of specifications over the three quantities $[X_1, X_2, X_3]$, and if we knew that $X_3 = X_1 + X_2$, and had specified variances coherently, then there would be at most two different axes of variation over $[X]$.

analagously with that for individual elements as the ratio of resolved to initial uncertainty. Hence our initial, adjusted and resolved uncertainties for the entire belief structure are:

$$
\begin{aligned}
\mathrm{U}(B) &= 2, \\
\mathrm{U}_D(B) &= 1.6614, \\
\mathrm{RU}_D(B) &= 0.3386
\end{aligned}
$$

The most useful result is the resolution of the belief structure:

$$\mathrm{R}_D(B) = 0.1693$$

showing that uncertainty over the entire belief structure $[B]$ is only reduced by about 17%.

4.4 Canonical directions

For a general belief adjustment we can construct a series of linear combinations, or **canonical directions**, with the properties that the first canonical direction is the linear combination of B_i's with the largest possible resolution; the second canonical direction is the linear combination with the largest resolution amongst those combinations which are uncorrelated with the first direction; and so forth. We take each canonical direction to have prior expectation zero and prior variance unity by convention. Our term for this collection of canonical directions is **belief grid**, as they comprise a multidimensional grid of directions over which the implications of the adjustment for the belief structure $[B]$ may be summarised as follows.

Each quantity in $[B]$ can be written as a linear combination of the canonical directions, and the resolution of each such quantity can be decomposed into a weighted sum of the resolutions for the canonical directions. The weights correspond to the strength of (squared) correlation between the quantity and the canonical directions. In this way, we expect to learn most about those quantities that are strongly correlated with the first few canonical directions; and least about those quantities that are weakly correlated with the first few canonical directions, and strongly correlated with the last few directions.

The belief grid for our simple example consists of the following two canonical directions:

$$
\begin{aligned}
Y_1 &= 1.01B_1 - 0.11B_2 - 3.47 \\
Y_2 &= -0.30B_1 + 0.70B_2 - 3.14
\end{aligned}
$$

or, in standardised form,

$$
\begin{aligned}
Y_1 &= 1.06B_1^* - 0.18B_2^* \\
Y_2 &= -0.32B_1^* + 1.10B_2^*
\end{aligned}
$$

$$\text{where } B_i^* = \mathrm{S}(B_i) = \frac{B_i - \mathrm{E}(B_i)}{\sqrt{\mathrm{Var}(B_i)}}.$$

with approximate resolutions

$$
\begin{aligned}
\mathrm{R}_D(Y_1) &= 0.32, \\
\mathrm{R}_D(Y_2) &= 0.02.
\end{aligned}
$$

Thus, the linear combination of quantities in $\langle B\rangle$ about which we expect to learn most is Y_1, and we expect to remove about 32% of our uncertainty in this direction. Any other linear combination of elements in $\langle B\rangle$ which is highly correlated with Y_1 will likewise have a similar variance reduction.

There is only one direction in $\langle B\rangle$ which is orthogonal to Y_1: we expect to learn least about Y_2. A resolution of only about 2% suggests that we learn almost nothing about both Y_2 and linear combinations highly correlated with Y_2.

Thus, for the purpose of learning about B_1 and B_2, the information contained in $[D]$ is essentially one-dimensional: we reduce uncertainty only in the direction of Y_1. Examination of the standardised form of the first canonical direction Y_1 above shows that B_1 is the major component, whereas B_2 is the major component of Y_2. Hence, we are learning mostly in the direction of B_1, and learning very little in the direction of B_2.

4.5 Adjusting the canonical belief structure

If we wish, we can define a new base $Y = \{Y_1, Y_2\}$ (where Y_1 and Y_2 are the canonical directions given in the preceding section) to contain our belief grid, and examine the adjustment of the corresponding belief structure $[Y]$ by the belief structure $[D]$. Amongst other results, we find that the summary of uncertainty over the belief structure $[Y]$ is the same as that for the belief structure $[B]$:

$$\begin{aligned} \mathrm{U}(Y) &= 2 = \mathrm{U}(B), \\ \mathrm{RU}_D(Y) &= 0.1693 = \mathrm{RU}_D(B). \end{aligned}$$

This is because the elements $\{Y_1, Y_2\}$ constitute a basis for the vector space $\langle B \rangle$, so that the corresponding belief structures $[B]$ and $[Y]$ are equivalent for this purpose. We also obtain the standardised adjusted expectation formula for each direction, approximately

$$\begin{aligned} \mathrm{E}_D(Y_1) &= 0.60D_1^* - 0.10D_2^*, \\ \mathrm{E}_D(Y_2) &= -0.05D_1^* + 0.16D_2^*. \end{aligned}$$

A further property here is that just as the original directions Y_i are uncorrelated a priori, so too the adjusted expectations for the canonical directions, $\mathrm{E}_D(Y_i)$ are uncorrelated.

5 Observing the adjustment

5.1 Standardised observations

An obviously useful check is to examine whether the data are consistent with the beliefs expressed about them. The simplest way of doing this is to standardise the data. Our notation for the standardised value of a datum x is $\mathrm{S}(x)$. For the data we saw, $d_1 = 5.4$ and $d_2 = 9.8$,

$$\begin{aligned} \mathrm{S}(d_1) &= \frac{d_1 - \mathrm{E}(D_1)}{\sqrt{\mathrm{Var}(D_1)}} = 1.17, \\ \mathrm{S}(d_2) &= \frac{d_2 - \mathrm{E}(D_2)}{\sqrt{\mathrm{Var}(D_2)}} = 2.28. \end{aligned}$$

The observation d_2 might be a little suspect: the Doctor's two-hour reading is more than two standard deviations distant from her expectation. The fasting measurement d_1 is just over one standard deviation from her expectation.

5.2 Evaluating the adjusted expectation

For our actual observations we evaluate the adjusted expectations shown above, and so obtain

$$\begin{aligned} \mathrm{E}_d(B_1) &= 4.71 \\ \mathrm{E}_d(B_2) &= 6.91. \end{aligned}$$

Our notation here is that $\mathrm{E}_d(X)$ refers to our adjusted expectation for X in the light of the information contained in $[D]$, having observed $D = d$. From the point of view of the Doctor, her belief specifications and the observations $d_1 = 5.4$ and $d_2 = 9.8$ that she makes when she performs the ogt-test upon herself, are consistent with her revising her expectations upwards for both B_1 and B_2. In the case of the fasting blood-glucose measurement, the analysis shows a revision upwards from 4.16 to 4.71; and in the case of the following two-hour measurement, a revision upward from 6.25 to 6.91.

Informally, as a very rough guide to the locations of B_1 and B_2, we might decide to take intervals of about two standard deviations in either direction from the expectation as being fairly likely to contain the relevant locations. For the prior assessments we have approximately the intervals

$$
\begin{aligned}
B_1 &: 4.16 \pm 2\sqrt{1.12} &= (2.04, 6.28),\\
B_2 &: 6.25 \pm 2\sqrt{2.43} &= (3.13, 9.38).
\end{aligned}
$$

For the assessments after adjusting by $[D]$ we obtain the smaller intervals

$$
\begin{aligned}
B_1 &: 4.71 \pm 2\sqrt{0.77} &= (2.96, 6.46),\\
B_2 &: 6.91 \pm 2\sqrt{2.32} &= (3.86, 9.96).
\end{aligned}
$$

The adjusted expectation for the two-hour blood-glucose measurement is 6.91, implying that an *average healthy* elderly patient will have a two-hour reading on the borderline between being diagnosed as healthy and being diagnosed as having impaired glucose tolerance. Put another way, about half of the elderly will be misdiagnosed according to the threshholds set for this ogt-test. Additionally, the fasting blood-glucose level has adjusted expectation 4.71; suggesting that the elderly have a slightly higher fasting level than do the young. Consequently the analysis suggests that there are static differences between the blood-glucose levels for the young and the elderly; and dynamic differences between their abilities to cope with fluctuations in blood-glucose levels.

5.3 Standardised adjusted expectations

We standardise each adjusted expectation to have expectation zero and variance unity, and evaluate these standardised quantities as follows:

$$
\begin{aligned}
\mathrm{S}_d(B_1) &= \frac{\mathrm{E}_d(B_1) - \mathrm{E}(B_1)}{\sqrt{\mathrm{RVar}_D(B_1)}} = 0.9296,\\
\mathrm{S}_d(B_2) &= \frac{\mathrm{E}_d(B_2) - \mathrm{E}(B_2)}{\sqrt{\mathrm{RVar}_D(B_2)}} = 2.0117.
\end{aligned}
$$

How do we interpret such values? For B_1 the change in expectation is about 0.93 standard deviations upward, a change that doesn't trouble us greatly. However, the change for B_2 is fractionally over two standard deviations, and so is somewhat larger than we would have expected: for a reduction in variance of only about 5%, we saw a relatively large change in expectation.

These standardised values serve as diagnostic flags. Here we have seen a value larger than expected, and we should consider the following possibilities:

- the data are more variable than expected, as $\mathrm{E}_d(B_2)$ is further from $\mathrm{E}_D(B_2)$ than expected;
- the resolved variation $\mathrm{RVar}_D(B_2)$ is smaller than would be consistent with such a change in expectation, implying that the prior variance $\mathrm{Var}(B_2)$ is too tight.

5.4 Consistency checks for the canonical directions

We have determined $\mathrm{E}_D(B_1)$ and $\mathrm{E}_D(B_2)$, and we can similarly obtain the adjusted expectation for any linear combination of B_1 and B_2.[5] In particular, we can evaluate the adjusted expectation for each canonical direction and determine their corresponding standardised values:

$$
\begin{aligned}
\mathrm{E}_d(Y_1) &= 0.48\\
\mathrm{E}_d(Y_2) &= 0.30.
\end{aligned}
$$

In each case, these adjusted values are changes from prior expectations of zero. The standardised changes are

[5]A property of the adjusted expectation is that it is a linear operator, so that for any scalars a_1 and a_2 we have

$$\mathrm{E}_D(a_1 B_1 + a_2 B_2) = a_1 \mathrm{E}_D(B_1) + a_2 \mathrm{E}_D(B_2)$$

.

$$\begin{aligned}\mathrm{S}_d(Y_1) &= 0.84 \\ \mathrm{S}_d(Y_2) &= 2.12.\end{aligned}$$

Thus, there are changes over both directions of the belief grid, but the change for the second direction (about which we are reducing uncertainty hardly at all, because we have seen that its resolution is $\mathrm{R}_D(Y_2) = 0.02$) is somewhat larger than expected, and warns us of a possible complication that may need investigation.

Examination of the standardised adjusted expectation formula for Y_2 (shown in section 4.5) shows us that the more influential quantity is D_2. This serves as yet another warning that D_2 has some peculiar feature.

5.5 The size and bearing of the adjustment

The **bearing** is a quantity which summarises the actual effects of adjustment. Its principal properties are that

- it is the linear combination in $\langle B \rangle$ having the largest standardised squared change in expectation;
- the change in expectation for any quantity in $\langle B \rangle$ is equal to the prior covariance between the quantity and the bearing;

The bearing for this adjustment is the linear combination

$$\mathrm{Z}_d(B) = 0.39B_1 + 0.16B_2 - 2.60.$$

or, standardising the B_i's,

$$\mathrm{Z}_d(B) = 0.41B_1^* + 0.25B_2^*.$$

We noted above the property that the changes from prior expectation to adjusted expectation for quantities in B are equivalent to the covariances of the quantities with the bearing. For example, for B_1,

$$\begin{aligned}\mathrm{E}(B_1) - \mathrm{E}_d(B_1) &= \mathrm{Cov}(\mathrm{Z}_d(B), B_1) \\ &= \mathrm{Cov}(0.39B_1 + 0.16B_2 - 2.60, B_1) \\ &= 0.39\mathrm{Var}(B_1) + 0.16\mathrm{Cov}(B_1, B_2) \\ &= 0.55,\end{aligned}$$

can be seen to be the difference between its expectation $\mathrm{E}(B_1) = 4.16$ and its adjusted expectation $\mathrm{E}_d(B_1) = 4.71$. In this way, changes in expectation are expressible solely through a covariance with the bearing, and so the magnitude of a change in expectation for any quantity depends only upon the strength of correlation between the quantity and the bearing, and upon the length (i.e. the variance) of the bearing.

The variance of the bearing is a quantity that we refer to as the **size of the adjustment**. Here it is

$$\mathrm{Size}_d(B) = \mathrm{Var}(\mathrm{Z}_d(B)) = 0.3179.$$

We define the **expected size of the adjustment** to be the prior expectation (where we replace $\mathrm{Z}_d(B)$ by $\mathrm{Z}_D(B)$) for the size of the adjustment. For our example, it is

$$\mathrm{E}(\mathrm{Size}_D(B)) = 0.3386.$$

We can show that the size of the adjustment always turns out to be equal to the resolved uncertainty for the belief structure, $\mathrm{RU}_D(B)$, as seen in section 4.3. In this sense we have that the expected magnitude of change in expectation is equal to the expected reduction in uncertainty over the entire base.

The ratio of the size of the adjustment to its expected size is termed the **size ratio**. Its value here is

$$\mathrm{Sr}_d(B) = \frac{\mathrm{Size}_d(B)}{\mathrm{E}(\mathrm{Size}_D(B))} = 0.9389.$$

We interpret the three size statistics as follows. The size of the adjustment (here equal to 0.3179) summarises the magnitude of actual change over the entire belief structure $[B]$. Relative to its prior variance, *every* linear combination in $\langle B \rangle$ has an actual squared change in expectation no greater than this value. For example, we can be sure that the actual change in expectation in the quantity $B^* = B_1 + B_2$ is bounded according to

$$(\mathrm{E}_d(B^*) - \mathrm{E}(B^*))^2 \leq 0.3179 \mathrm{Var}(B^*) = (1.26)^2.$$

where $\mathrm{Var}(B^*) = 4.99$. (In fact the actual change in expectation turns out to be rather close to this bound, being $\mathrm{E}_d(B^*) = 1.21$.) We contrast the value for the size of the adjustment to its expectation by calculating their ratio, the **size ratio**. For our example the value is 0.9389, very close to its expectation of unity. Generally, the magnitude of the size ratio serves as a further diagnostic test: a size ratio distant from unity may cause us to reconsider our specifications, or to check our data values for spurious entries, and so forth.

6 Adjusting beliefs in stages

In our example so far we have adjusted one belief structure, $[B]$, by another, $[D]$. Have we exhausted our exploratory possibilities, or are there extra insights to be had by approaching the problem in a different way? Well, suppose that we consider the analogy of a traditional multiple regression, where we try to predict a response variable Y from a collection of regressors $X_1, \ldots, X_k$. In terms of this analogy we may be interested not only in the predictive power of the collection taken as a whole but also in whether *every* X_i is useful for the prediction; whether certain *subsets* of the X_i's are more useful than others; and so forth.

For our Bayes linear approach we are concerned with similar (but broader) interests, which we investigate by applying the notion of adjusting beliefs in stages. An illustration arises from queries raised about our example so far: two of our analyses to date (the standardised values for the adjusted expectations of both the original quantities and the canonical directions) suggest actual changes in expectation substantially larger than expected. Simultaneously we expect to learn very little about B_2 (as its resolution is only about 5%) and Y_2 (its resolution is only some 2%), but their changes in expectation are relatively very large. Various evidence points to a surprisingly large value of d_2 being at fault. Suppose, then, that we consider D_1 and D_2 as being two distinct sources of information, and suppose that we adjust $[B]$ firstly by D_1, and then by D_2. (In what follows we use the notation $[D]$ and $[D_1 \cup D_2]$ synonymously.)

6.1 Performing the initial adjustment

If we adjust $[B]$ solely by the Doctor's fasting measurement D_1, we find that as a source of information D_1 is almost as effectual as D_1 and D_2 combined. For example the resolution for $[B]$ is $\mathrm{R}_{D_1}(B) = 0.1554$, whereas we had earlier $\mathrm{R}_D(B_1) = 0.1693$. Our adjusted expectations are

$$\begin{aligned} \mathrm{E}_{d_1}(B_1) &= 4.85 \\ \mathrm{E}_{d_1}(B_2) &= 6.58 \end{aligned}$$

showing relatively small (1.17 standard deviations) changes from her initial expectations. Her adjusted variances are approximately

$$\begin{aligned} \mathrm{Var}_{D_1}(B_1) &= 0.78 \\ \mathrm{Var}_{D_1}(B_2) &= 2.35 \end{aligned}$$

which are only slightly larger than those shown in section 4.2 for the full adjustment. We see also a size ratio of

$$\mathrm{Sr}_{d_1}(B) = 1.3729,$$

implying no conflict between data and prior belief as sources of information in the sense that this value is close to unity, its expectation.

6.2 Partial adjustments of belief

From the point of view of variance reduction, the adjustment of the B_i's solely by D_1 seems very little different to that obtained by adjusting fully by D_1 and D_2. We can examine this suggestion, and the determine the effects upon expectation, by considering separately the effects of adjusting by the two-hour measurement, D_2, in addition to D_1.

6.2.1 Partial variances

Every additional adjustment has the potential to reduce further the uncertainty in our quantities of interest. The extra variance reductions due to these partial adjustments are called partial resolved variances. When we adjust by D_2 in addition to D_1, the extra reductions in variance for B_1 and B_2 are as follows:

$$\begin{aligned} \mathrm{RVar}_{[D_2/D_1]}(B_1) &= 0.0049 \\ \mathrm{RVar}_{[D_2/D_1]}(B_2) &= 0.0286 \end{aligned}$$

with resolutions

$$\begin{aligned} \mathrm{R}_{[D_2/D_1]}(B_1) &= 0.0044 \\ \mathrm{R}_{[D_2/D_1]}(B_2) &= 0.0118. \end{aligned}$$

Thus, the partial effect of adjusting by D_2 additionally is negligible: relative to the initial uncertainty in B_2, we achieve a further reduction in uncertainty of only some 1.18%, and the relative reduction in variance for B_1 is smaller still.

We may also evaluate the reductions in uncertainty relative to the adjustment variances for the current adjustment:[6]

$$\begin{aligned} \mathrm{R}_{[D_2/D_1]}([B_1/D_1]) &= 0.0064, \\ \mathrm{R}_{[D_2/D_1]}([B_2/D_1]) &= 0.0122, \\ \mathrm{R}_{[D_2/D_1]}([B/D]) &= 0.0164. \end{aligned}$$

Thus, relative to the adjusted variances calculated for the initial adjustment, $\mathrm{Var}_{D_1}(\cdot)$, we see reductions of only 0.64% and 1.22% for B_1 and B_2 respectively; and relative to the former adjustment uncertainty $\mathrm{U}_{D_1}(B)$ a further reduction in uncertainty for the overall belief structure $[B]$ of only 1.64%. The conclusion seems quite clear: from the point of view of reducing uncertainties, D_2 as an additional source of information is almost worthless; D_1 alone carries most of the information.

The notation that we use here, $\mathrm{R}_{[D_2/D_1]}([B_1/D_1])$ for example, reflects the fact that after the initial adjustment by D_1, it is as though we focus attention on the portion of B_1 that remains uncertain, namely $[B_1/D_1]$, the adjusted version of B_1 given D_1. Now a partial adjustment by $[D_2/D_1]$ of the original quantity B_1 is equivalent to an initial adjustment of the adjusted version (which has prior variance $\mathrm{Var}([B_1/D_1]) = \mathrm{Var}_{D_1}(B_1)$) by $[D_2/D_1]$.

6.2.2 Partial canonical directions

The solitary partial canonical direction is:

$$W = 0.80B_1 - 0.65B_2 + 0.77,$$

with an extra partial resolution of $\mathrm{R}_{[D_2/D_1]}(W) = 0.0277$. These results show that one effect of the partial adjustment will be to reduce uncertainty in the overall belief structure $[B]$ by only 2.8%, and that this is the maximum partial resolution for any linear combination in $\langle B \rangle$. The degree of resolution for any such linear combination depends upon the strength of correlation between it and the partial canonical direction W.

[6] When it is the case that the current adjustment variance is already a small fraction of the initial variance, a further variance reduction might represent only a small partial resolution. However, this might also represent a substantial reduction relative to the current adjustment variance.

6.2.3 Partial expectations

The effect of the partial adjustment by D_2 on the evaluation of the adjusted expectation is to revise expectations downwards for B_1, from about 4.85 to 4.71, and upwards for B_2 from about 6.58 to 6.91. (These are now, of course, the adjusted expectations for the full adjustment by $[D]$.) The differences are, in our notation,

$$\begin{aligned}\mathrm{E}_{[D_2/D_1]}(B_1) &= \mathrm{E}_D(B_1) - \mathrm{E}_{D_1}(B_1) = -0.14\\ \mathrm{E}_{[D_2/D_1]}(B_2) &= \mathrm{E}_D(B_2) - \mathrm{E}_{D_1}(B_2) = +0.33.\end{aligned}$$

Standardising the changes, we have

$$\mathrm{S}_{[d_2/d_1]}(B_1) = \mathrm{S}_{[d_2/d_1]}(B_2) = 1.96,$$

representing, for each quantity, a change in expectation of 1.96 standard deviations relative to the partial variance resolved. That is, for such a small change in variance we saw a fairly large change in expectation by using the observed two-hour measurement as well as the observed fasting measurement.

6.2.4 The size of the partial adjustment

In the same way that we can evaluate a size and an expected size for any general adjustment, we can also evaluate a size and an expected size for a partial adjustment. For our example, the sizes and size ratios for the full adjustment; the partial adjustment; and the simple adjustment by D_1 only are shown in figure 2.

Figure 2: Sizes for the adjustments D_1 and $[D_2/D_1]$

	Adjustment		
	$D = [D_1 \cup D_2]$	$[D_2/D_1]$	D_1
Size	0.3179	0.1069	0.4268
Expected	0.3386	0.0277	0.3109
Size ratio	0.9389	3.8524	1.3729

The size ratio for the overall adjustment is roughly $\mathrm{Sr}_d(B) = 0.94$, and the size ratio for the simple adjustment by D_1 is $\mathrm{Sr}_{d_1}(B) = 1.37$, neither being surprising values. However the size ratio for the partial adjustment is $\mathrm{Sr}_{[d_2/d_1]}(B) = 3.85$, nearly four times more than expected and suggesting an inconsistency between prior specification and data for the additional adjustment by D_2.

6.2.5 The bearing and path correlation for the partial adjustment

We have seen already the bearing for the full adjustment (in section 5.5) and we can calculate the bearing for the adjustment solely on D_1, giving a vector essentially in the direction of B_1:

$$\mathrm{Z}_{d_1}(B) = 0.65B_1 - 0.06B_2 - 2.35.$$

The difference between the two is the bearing for the partial adjustment:

$$\mathrm{Z}_{[d_2/d_1]}(B) = -0.26B_1 + 0.21B_2 - 0.25.$$

It is solely in this direction that expectations can change according to the new information contained in D_2. Clearly these two directions (which aggregate to form the overall bearing) are different, so that the partial adjustment is telling us something extra about a different direction. We can summarise this by evaluating the **path correlation**: the prior correlation between the previous and partial adjustment bearings. Here the path correlation is

$$\mathrm{C}(d_1, [d_2/d_1]) = -0.5051,$$

showing that from the point of view of revising expectations, the data are partly contradictory. Had this correlation been positive, we would have argued that the data complemented each other, with the magnitude of correlation indicating the degree of consistency. Our interpretation follows from our being able to write

$$\mathrm{Size}_d(B) = \mathrm{Size}_{d_1}(B) + \mathrm{Size}_{[d_2/d_1]}(B) + 2\mathrm{Cov}(\mathrm{Z}_{d_1}(B), \mathrm{Z}_{[d_2/d_1]}(B)).$$

Thus, it is the negative covariance (which we see summarised as a path correlation of $\mathrm{C}(d_1, [d_2/d_1]) = -0.5051$) between the bearings for the previous and partial adjustments which serves to diminish the size of the joint adjustment: the changes in expectation for the initial and partial adjustments are in different directions, and thus tend to cancel out each other.

6.3 Withdrawing quantities from the adjustment

In the same way that we can introduce additional quantities into the adjustment, so too can we determine the effects of withdrawing quantities from the adjustment. We might do this for various reasons,[7] but here particularly for the purpose of investigating the rather peculiar nature of the specifications over D_2 and its observed value d_2.

When we remove D_1 from the adjustment at this stage, it as though we are left with a simple adjustment of $[B]$ by D_2. In addition, we also learn about the partial adjustment of $[B]$ by $[D_1/D_2]$. Some of the results of this partial adjustment are worth further consideration as follows. The resolutions for the adjustment by D_2 and for the partial adjustment removing D_1 are:

$$\begin{aligned} \mathrm{R}_{D_2}(B) &= 0.0448 \\ \mathrm{R}_{[D_1/D_2]}(B) &= 0.2938 \end{aligned}$$

showing - as we suspected - that D_2 alone is not a good source of information: its effect is at best to reduce uncertainty by less than 5%; and that most of the information is contained wholly in D_1: when we remove $[D_1/D_2]$ we also remove nearly all of our capability to reduce uncertainty in $[B]$.

The evaluated adjusted expectations for the adjustment by D_2 solely are

$$\begin{aligned} \mathrm{E}_{d_2}(B_1) &= 4.60 \\ \mathrm{E}_{d_2}(B_2) &= 6.88, \end{aligned}$$

representing fairly large changes ($\mathrm{S}_{d_2}(B_1) = \mathrm{S}_{d_2}(B_2) = 2.3$ standard deviations) from the initial values of 4.16 and 6.25 respectively. Although these are fairly large changes, they correspond to only small reductions in uncertainty about B_1 and B_2. Informally it is as though the prior and posterior distributions for B_1 and B_2 are quite flat, so that large changes in expectation can correspond to small changes in uncertainty.

Figure 3: Sizes for the adjustments D_2 and $[D_1/D_2]$

	Adjustment		
	$D = [D_1 \cup D_2]$	$[D_1/D_2]$	D_2
Size	0.3179	0.0115	0.2325
Expected	0.3386	0.2938	0.0448
Size ratio	0.9389	0.0390	5.1862

The sizes of the adjustments are shown in figure 3 and display two remarkable features. Firstly, the size ratio for the simple adjustment by D_2 is more than five times as large as expected; and secondly the size ratio for the partial adjustment after adjusting by D_1 additionally is very much smaller than expected. The former feature is more or less expected, given the sizes for the similar adjustment shown in figure 2. The latter feature may be interpreted as showing that a partial adjustment by D_1 in addition to D_2 is expected to enable changes in expectation that simply don't materialise.

[7]For example, we might remove uninformative quantities, or quantities that are relatively unimportant and expensive to observe.

Comparing figure 3 with figure 2, we see a size ratio of $\mathrm{Sr}_{d_2}(B) = 5.1862$ for the simple D_2 adjustment, and a size ratio of $\mathrm{Sr}_{[d_2/d_1]}(B) = 3.8524$ for the partial adjustment by $[D_2/D_1]$. Thus, the size ratio for the simple D_2 adjustment is larger than the size ratio for the adjustment where D_1 has been extracted. This might suggest that although we have identified D_2 (with its observation d_2) above as having some peculiar features, this is also true of the portion of D_1 that is common to D_2.

7 Review of the example

We began with quantities B_1 and B_2, quantities representing the blood glucose levels before, and two hours after, ingestion of a certain quantity of sugar for an "average" or "typical" elderly person. To learn about these quantities, an elderly doctor takes the similar quantities D_1 and D_2 relating specifically to herself, and makes various specifications (expectations and covariances) over these quantities; and observes D_1 and D_2.

Our approach shows that the following "answers" are consistent with the doctor's inputs.

- she revises her expectation for B_1 upwards from 4.16 to 4.71; and for B_2 upwards from 6.25 to 6.91. After standardising, we notice that the latter change is surprisingly large. We also notice that the adjusted expectation for B_2 is very close to a diagnosis threshhold.
- she reduces her variance for B_1 from 1.12 to 0.77, and for B_2 from 2.43 to 2.32, reductions of some 31% and 5% respectively.
- she examinines the canonical directions and discovers that she is learning mostly in only one direction, the first canonical direction $Y_1 = 1.01B_1 - 0.11B_2 - 3.47$, suggesting that if she wished to improve her estimates, she should concentrate on learning about the other canonical direction, $Y_2 = 0.7B_2 - 0.3B_1 - 3.14$, perhaps by considering further observables that she feels might be informative about the speed of reaction of an elderly person's body to various stimuli.
- the standardised adjusted expectation formula for Y_2 shows that D_2 is the most influential component of Y_2. Thus she checks her data values and finds that the standardised value for d_2 is much larger than its expectation.
- by examining the bearing, she discovers the most important direction for actual adjustment, and she observes that the size ratio is 0.9389, suggesting that the influence of the data for the change in expectation was about as expected.
- some of the analyses suggest actual changes in expectation significantly larger than expected. We explored our data and specifications further by adjusting in stages, and found two points of interest:
  - the two-hour measurement D_2 is not expected to be informative in addition to what can be learned from D_1, so that $[D_2/D_1]$ is essentially non-informative;
  - it is precisely this portion that produces surprising changes in expectation given its lack of information content.
  - these conclusions can be drawn also from withdrawing D_1 from the adjustment and examining the effects upon the inflation of uncertainties and the changes in expectation.

Without exploring the inputs further, one or both of the following interpretations appear plausible. Firstly, our Doctor may not be as typical a healthy elderly patient as she believes herself to be: this interpretation is supported by her fasting measurement also being rather larger than expected. Secondly, she may be more certain about her expectations than she should be. If she is sure that she is healthy, then her hypothesis that the elderly are misdiagnosed is supported by this analysis; however she has allowed insufficient leeway in the variance specifications for the data to be consistent with her beliefs. In a sense she is "more right" about her doubts about the ogt-test than she suspected.

Finally, whereas the analysis has suggested these conclusions as being consistent with the Doctor's inputs, neither the inputs nor the conclusions are definitive for various reasons. For example,

- she might have provided more detailed inputs if she had been prepared to put in the extra work necessary beforehand;

- the analysis leaves open some questions;
- the analysis suggests that some of her inputs might be untrustworthy.

Therefore, she takes the adjustments of belief provided by the analysis as inputs helpful in forming her actual revised opinions. For example, the analysis has revealed that she is learning very little about D_2, so a natural continuation of her analysis would be to explore ways (perhaps a more detailed prior specification) of increasing her knowledge about D_2.

On another level, several technical issues are raised. For example, having performed the analysis, her judgement about whether or not she actually believes in the results of her analysis is itself a further statement of belief, itself subject to the laws of coherence over time. Such matters, although outside the scope of this document, are discussed in [1].

Part III
An introduction to the [B/D] language

In this part of the document we introduce the programming language [B/D] by using it to produce the results needed for the analysis given in the second part for the oral glucose tolerance test example. We concentrate on introducing the [B/D] commands necessary for the analysis, and on stating what the output represents technically. We do not restate any of the interpretation and analysis given earlier, but we will refer you back to the second part for such interpretation.

8 Introducing [B/D]

Here we begin our first tutorial in the [B/D] programming language, our computer implementation of the Bayes linear methodology. We begin by showing you how to start the program, and how to stop! Once we are underway, we introduce some simple, but basic concepts: the declaration of quantities and collections of quantities; the specification of covariances between quantities; and observations made on quantities. Throughout these early tutorial sessions we will try and keep our usage and description of the program as simple as possible - this is not the place to describe the much richer set of possibilies that the [B/D] language offers.

8.1 Starting [B/D]

We are going to run [B/D] interactively, simply by typing "bd" at the main system prompt (DOS or UNIX version), or by running the program **bd.exe** from Windows 3.1. After a few seconds, the program will load and will print a copyright message similar to that shown in figure 4, together with an indication of available memory and whether or not a maths co-processor is being used.

Figure 4: [B/D] start-up message

```
[B/D] V:8.14   (August 1995)      (C) David Wooff & Michael Goldstein
                                          University of Durham
```

When the copyright message has printed, you will see a prompt such as "BD>", which means that [B/D] is awaiting a command from you. Now we are ready to begin.

8.2 [B/D] Command Lines

Before we start pressing buttons, a word on the style that we have adopted. Each [B/D] command line consists of a command followed by a list of arguments to the command. The case (upper or lower) that you use is not important: the text that you input is generally converted internally to lower case, with some minor exceptions to handle headings of output and similar titling. The command must be separated from its arguments by at least one space, or by one colon. The letters forming the name of a command must be contiguous, although you can intersperse spaces freely in the argument list. Each line is terminated by a carriage return which we indicate with the ↩ symbol.

We are going to use a **bold** typeface to show the text that you should type at the computer keyboard, and we will usually preface this text with a prompt (such as `BD>`) which you should not reproduce. Remember to complete each line with a carriage return.

As an example, we will issue the command that terminates the program. This is the <u>STOP:</u> command, which is our usual way of ending a [B/D] session. Try it now by typing the command below, and then restart the program.

`BD>`**stop:** ↩

When the program has restarted, enter the following command:

BD>**keep:log=mylog** ↩

This asks [B/D] to maintain a record of the results output during the session in the file called "mylog" in your current directory. (Any file of the same name that already exists there will be overwritten.) You may choose a file name other than "mylog" (subject to the file-naming convention for your computer) if you wish.

9 Organising inputs to [B/D]

9.1 Declaring quantities

In our first example we are interested in the quantities B_1 and B_2, and we intend observing further quantities D_1 and D_2 to help us refine our belief specifications about the former collection. The first thing that we must do is to announce to [B/D] these quantities of interest. We can do this, and at the same time specify our expectations for the quantities, by using the [B/D] command ELEMENT: as follows:

BD>**element:B1=4.16, B2=6.25, D1=4.16, D2=6.25** ↩

If you entered this line correctly, [B/D] will have noted B1, B2, D1, and D2 as quantities, and will have assigned the appropriate expectation to each. The names that we use in [B/D] (we use B1 for the quantity B_1, for example) are chosen to correspond informally with the names of actual quantities. For genuine problems we normally use longer and more informative names, such as "glucoselevel0", although there are restrictions[8] upon the maximum length of such names. The names of quantities consist of mixtures of alphabetic and numeric characters, and must begin with a letter[9]. Remember that both the case and interspersed spaces are irrelevant, so that the names "PEPTIDE 17" and "peptide17" are equivalent.

You might have entered the line wrongly, either by mistaking the syntax of the command or by mistyping the names or expectations of the quantities. In the former case [B/D] will have reported your syntactical error, and you can simply retype the line. This is what we usually do when [B/D] reports an error. In the latter case (harder to spot because [B/D] has accepted the line as valid input) you can either retype the whole line, or just retype the names and expectations that you had wrong. If we wish, we can examine our specifications by issuing the command

BD>**look:(e)** ↩

Figure 5: Elements and their expectations

Element	Expectation
B1	4.160000
B2	6.250000
D1	4.160000
D2	6.250000

giving the output shown in figure 5, which shows all elements defined, together with their expectations. This is the first of many uses that we shall make of the LOOK: command, which, with its variety of possible arguments, will form our main tool for checking and reviewing our input to the program.[10]

9.2 Collections of quantites

We will find it helpful later if we organise our quantities into the natural collections of interest. In our example our four quantities are $\{B_1, B_2\}$, the collection of quantities about which we wish to learn; and $\{D_1, D_2\}$, the collection of observable quantities. Our term for collections like these is **base**, and within [B/D] we define these structures by using the BASE: command as follows:

[8] There are various restrictions like this which vary from machine to machine. The principal restrictions can be found by issuing the LOOK:(program) command.

[9] Many of the punctuation symbols have a special meaning in [B/D].

[10] A summary of the possible arguments to the LOOK: command, together with their meanings, can be seen by issuing the command with a question mark as argument: LOOK:(?).

BD>**base:B = B1, B2** ↩
BD>**base:D = D1, D2** ↩
BD>**base:G = B, D** ↩

Here we have specified the two natural bases, named "B" and "D" (there is no extra meaning attached to using the names B and D; we could have used "bacon" and "eggs", or anything else that takes our fancy) and a further base named "G" to contain all the quantities for our later convenience. Notice that we defined this last base in terms of previously defined bases[11]. The names that we use are subject to the same rules that are used to name elements. We can check our definitions by issuing the command

BD>**look:(b)** ↩

Figure 6: Bases defined

```
Base
B    :  B1, B2
D    :  D1, D2
G    :  B, D
```

The output, shown in figure 6, lists the names of the bases that [B/D] knows about, along with their contents. Notice that the component quantities of the bases are listed in alphabetical order. The ordering is important because we will frequently use a base name as shorthand for the collection of quantities that it represents, and we will need to know the order in which quantities within a base will be affected by our actions. Notice also that the base "G" contains other bases rather than elements.

9.3 Specifying beliefs

We have introduced to [B/D] our four elements, and we have specified expectations for them, but we have yet to specify variances and covariances for them. In fact [B/D] assumes that all variance and covariance specifications are zero until informed otherwise. This we do by using the [B/D] command VAR:. Let us look at a simple example of the use of this command:

BD>**var:v(1,B1,B2) = 0.72** ↩

This syntax stands for "take the covariance between B1 and B2 in store 1 to be 0.72". Throughout this document we will ignore the store number as it will always be 1[12].

In our example with four quantities we have ten such beliefs to announce: four variances and six covariances. We could enter each of these ten beliefs individually into [B/D] as in the example above, but instead we will save some effort by declaring beliefs over the entire collection at the same time. We do this as follows. The base named "G" can be thought of as a vector:

$$G = \begin{bmatrix} B \\ D \end{bmatrix} = \begin{bmatrix} B_1 \\ B_2 \\ D_1 \\ D_2 \end{bmatrix},$$

Here, the base "G" is replaced by its constituent elements, in alphabetical order. Now think of the covariance matrix over the vector G:

$$\mathrm{Var}(G) \quad = \quad \begin{bmatrix} \mathrm{Var}(B_1) & \mathrm{Cov}(B_1,B_2) & \mathrm{Cov}(B_1,D_1) & \mathrm{Cov}(B_1,D_2) \\ \mathrm{Cov}(B_1,B_2) & \mathrm{Var}(B_2) & \mathrm{Cov}(B_2,D_1) & \mathrm{Cov}(B_2,D_2) \\ \mathrm{Cov}(B_1,D_1) & \mathrm{Cov}(B_2,D_1) & \mathrm{Var}(D_1) & \mathrm{Cov}(D_1,D_2) \\ \mathrm{Cov}(B_1,D_2) & \mathrm{Cov}(B_2,D_2) & \mathrm{Cov}(D_1,D_2) & \mathrm{Var}(D_2) \end{bmatrix}$$

[11] We can use wildcard facilities to generate complicated bases quickly. In particular, the symbol '$' after a name means all those elements whose names match as far as the '$' symbol; for example we could have used "B$" as shorthand for "B1, B2" above.
[12] [B/D] allows several different belief stores, indexed by store numbers, intended for alternative belief specifications.

$$= \begin{bmatrix} 1.12 & 0.72 & 0.62 & 0.30 \\ 0.72 & 2.43 & 0.30 & 0.43 \\ 0.62 & 0.30 & 1.12 & 0.72 \\ 0.30 & 0.43 & 0.72 & 2.43 \end{bmatrix}.$$

Now we input the lower triangle of this covariance matrix, where each row of numbers must be entered before pressing the carriage return key:

`BD>`**var:v(1,G)** ↩
`BD*`**1.12** ↩
`BD*`**0.72 2.43** ↩
`BD*`**0.62 0.30 1.12** ↩
`BD*`**0.30 0.43 0.72 2.43** ↩

Hence, this form of the VAR: command allows us to input the lower triangle of the variance matrix for the collection defined by the name "G", and in so doing defines all the necessary covariances. There are other styles of introducing covariance specifications, for example we might well have introduced $\mathrm{Var}(B)$, $\mathrm{Var}(D)$, and $\mathrm{Cov}(B, D)$ separately.

Did you notice that the [B/D] prompt changed from `BD>` to `BD*` after you typed in the first line? This is because [B/D] has worked out that you need to input four rows of a lower triangular matrix, and so expects strictly numerical input, where the numbers are separated by at least one space. You should type in a complete row of numbers before ending the line by typing a carriage return. The prompt will revert to the standard prompt, `BD>`, as soon as the numerical input is complete. ([B/D] uses different prompts to indicate the kind of input that it expects and its type of environment.)

We can check our specifications by issuing the command

`BD>`**look:(v1)** ↩

Figure 7: Covariance definitions

```
Covariances in store (1) :
B1 :     1.1200    0.7200    0.6200    0.3000
B2 :     0.7200    2.4300    0.3000    0.4300
D1 :     0.6200    0.3000    1.1200    0.7200
D2 :     0.3000    0.4300    0.7200    2.4300
```

giving the output shown in figure 7, which lists the belief specifications for all elements so far defined. The number that you specified for $\mathrm{Cov}(B_1, B_2)$ in the first simple example of the VAR: command that you typed in has been overwritten; if you notice any mistakes, you can correct them piecemeal by using this single-entry syntax.

By replacing the `v1` argument by the `r1` argument, we get to see the correlation matrix for our quantities of interest, rather than the covariance matrix. The output shown in figure 8 is obtained by issuing the command:

`BD>`**look:(r1)** ↩

Figure 8: Correlations

```
Correlations in store (1) :
B1 :     1.0000    0.4364    0.5536    0.1818
B2 :     0.4364    1.0000    0.1818    0.1770
D1 :     0.5536    0.1818    1.0000    0.4364
D2 :     0.1818    0.1770    0.4364    1.0000
```

9.4 Inputting data

We declare actual observations by using the DATA: command. In our simple example there are but two pieces of data (observations of $d_1 = 5.4$ and $d_2 = 9.8$) which we declare to [B/D] as follows:

BD>**data:D1(1) = 5.4** ↩
BD>**data:D2(1) = 9.8** ↩

Here, the notation D1(1) simply means the first observation on D1. Had we more than one observation, then we would be defining D1(2), D1(3), and so on. You can think of the data as being stored in a large matrix where the columns represent different quantities, and the rows represent different cases. We generally refer to the columns as data vectors. (The quantities need not have the same number of observations, and observations can be "missing", but you might need to be careful when mixing such data vectors.) We can examine our data specifications by issuing the command

BD>**look:(d)** ↩

Figure 9: Data definitions

```
Data storage   :
Case         d1          d2
  1       5.400       9.800
```

giving the output shown in figure 9, listing all the elements for which data has been specified, together with their data values.

10 Adjusting beliefs by data

10.1 Performing an adjustment

In our first example we obtain the adjustment of the belief structure $[B] = [B_1, B_2]$ by belief structure $[D] = [D_1, D_2]$. Here is the [B/D] command which determines the adjustment (remember that the quantities B and D in the command are bases which we have already defined[13]):

BD>**adjust:[B/D]** ↩

No output resulted when you typed this command (unless you made a mistake, in which case you may re-enter the command). This is because [B/D] only reports results that we explicitly request, and whilst we could have requested various results beforehand (but didn't), we will find it more convenient here to use the SHOW: command as in the following sections.

The ADJUST: command is of key importance within the program, and there are a great many program switches which can influence what happens precisely when you issue such a command. For our purposes, the default configuration[14] is suitable.

The variance and covariance specifications made in section 3.2 are *coherent* in that the variance-covariance matrix for the vector $G = [B_1\, B_2\, D_1\, D_2]^T$ is non-negative definite. Such non-negative definiteness is a requirement for our approach (that is, we disallow the possibility of negative variances). The ADJUST: command automatically checks that your various inputs are coherent, and advises you when they are not, by reporting the error and halting the adjustment.

Many of the effects of an ADJUST: command can be viewed by supplying various *options* to the SHOW: command. An *option* in [B/D] is used to specify the type and amount of output that we see. For many options, greater detail can be had by changing the pertinent output option slightly. In fact we add a '+' symbol to the

[13]Recall that "B" and "D" are the names of our two bases. Had they been called "bacon" and "eggs", we would have issued the command ADJUST:[bacon/eggs].

[14][B/D] has a default configuration which allows us to analyse straightforward problems without worrying too much about esoterica. In general, though, we have to consider the settings of the various [B/D] *controls* which control the action of the program. We will see some of these controls in use later.

option. To help you, we will show the names of options in a different typeface like this. Initially we look at three of the most basic types of output: adjusted expectations and evaluations of them, and adjusted variances.

10.2 Adjusted expectations

We can see the adjusted expectations by selecting the e option of the SHOW: command as follows:

BD>**show: e** ↩

Figure 10: Adjusted expectations

```
Element          Adjusted Expectation
B1 :         0.5858 D1  - 0.0501 D2  + 2.0363
B2 :         0.1904 D1  + 0.1206 D2  + 4.7047
```

yielding the output shown in figure 10. The output shows us the adjusted expectations $\mathrm{E}_D(B_1)$ and $\mathrm{E}_D(B_2)$ for the elements B_1 and B_2.

We can see the standardised adjusted expectations by selecting the e* option of the SHOW: command as follows:

BD>**show: e*** ↩

Figure 11: Standardised adjusted expectations

```
Element          Adjusted Expectation (standardized)
B1 :         0.6199 D1  - 0.0781 D2  + 4.1600
B2 :         0.2015 D1  + 0.1879 D2  + 6.2500
```

yielding the output shown in figure 11. The functions shown here are the same as those seen in figure 10, but with the data quantities D_1 and D_2 standardised to have expectation zero and standard deviation unity. For a discussion of all these adjusted expectations see section 4.1.

10.3 Adjusted variances

We display the adjusted variances by using the v option of the SHOW: command:

BD>**show: v** ↩

Figure 12: Adjusted variances

```
Element     Adjusted Variation
B1                 0.7718
B2                 2.3211
```

giving the output shown in figure 12. This shows the adjusted variance (informally the remaining uncertainty) for each quantity of interest: $\mathrm{Var}_D(B_1)$ and $\mathrm{Var}_D(B_2)$.

If we change our option from v to v+ we see further details on the adjusted variances:

BD>**show: v+** ↩

giving the output shown in figure 13. The results for both the simple and the detailed output are discussed in section 4.2. Technically, for individual quantities this output consists of the following information: reading from left to right (and using the output for B_1 to illustrate) we have

Figure 13: Adjusted variance details

	-------- Variation	-----------		Current
Element	Initial	Adjustment	Resolved	Resolution
B1	1.1200	0.7718	0.3482	0.3109
B2	2.4300	2.3211	0.1089	0.0448
[B]	2	1.6614	0.3386	0.1693

- the name of the element, B_1;
- the element's initial variance, $\mathrm{Var}(B_1)$;
- the element's adjusted variance, $\mathrm{Var}_D(B_1)$;
- the element's resolved variance, $\mathrm{RVar}_D(B_1)$;
- the element's current resolution, $\mathrm{R}_D(B_1)$.

The last row of the output (interpreted in section 4.3) contains results about the effect of adjustment on uncertainties within the entire belief structure $[B]$ taken as a whole. Reading from left to right, we have

- the name of the belief structure of interest, $[B]$;
- we summarise the prior uncertainty in the belief structure $[B]$ as the rank of the variance matrix specified over $[B]$, this rank being equal to the number of canonical directions for the belief structure. The notation we use for this summary of the uncertainty in the belief structure is $\mathrm{U}(B)$.
- we summarise the adjustment uncertainty of the belief structure by the sum of the adjustment variances for the canonical directions. Our notation for this quantity is $\mathrm{U}_D(B)$.
- we summarise the resolved uncertainty of the belief structure as the sum of the resolved variances for the canonical directions. Our notation for this is $\mathrm{RU}_D(B)$.
- Analagously with the individual elements, the resolution for the belief structure is the resolved uncertainty divided by the prior uncertainty. We will use the same notation for belief structures as well as for individual elements, and so our notation for this quantity is $\mathrm{R}_D(B)$.

10.4 Canonical directions

We can see the canonical directions of a particular adjustment by using the <u>cd+</u> option of the <u>SHOW:</u> command:

BD>**show: cd+** ↩

Figure 14: Belief grid

Canonical Directions :		
	1	2
Resolution	0.3184	0.0202
B1	1.0059	0.3020
B2	-0.1136	-0.7039
Constant	-3.4745	3.1431

giving the output shown in figure 14.[15] This shows two canonical directions, roughly $Y_1 = 1.01B_1 - 0.11B_2 - 3.47$ and $Y_2 = -0.30B_1 + 0.70B_2 - 3.14$, defined as linear combinations of the quantities of interest. These directions (which were discussed further in section 4.4) have prior covariance zero and prior variance unity, and are output together with their canonical resolutions, $\mathrm{R}_D(Y_1) = \mathrm{RVar}_D(Y_1)$ and $\mathrm{R}_D(Y_2) = \mathrm{RVar}_D(Y_2)$.

[15] We can issue the command without the "+" symbol, yielding only the canonical resolutions.

11 Observing the adjustment

11.1 Evaluating the adjusted expectation

We have shown what we expect to happen: let us now introduce data into the analysis by evaluating the adjusted expectations shown in figure 10 for our actual observations, $d_1 = 5.4$ and $d_2 = 9.8$. We can see this output if we select the a option of the SHOW: command:

BD>**show: a** ↩

Figure 15: Adjusted expectations evaluated

Element	Adjusted Expectation
B1	4.7085
B2	6.9140

giving the output shown in figure 15. The numbers given here are the adjusted expectations, $\mathrm{E}_d(B_1)$ and $\mathrm{E}_d(B_2)$, for the two quantities of interest, given the particular observed d.

As before, we can add the '+' symbol to our option to obtain further details on the evaluations of the adjusted expectations:

BD>**show: a+** ↩

Figure 16: Adjusted expectations details

	Expectations:		
Element	Initial	Adjusted	Std
B1	4.1600	4.7085	0.9296
B2	6.2500	6.9140	2.0117

giving the output shown in figure 16. This output (which is discussed further in section 5.2) gives the following information: reading from left to right, and using B_1 to illustrate, we have

- the name of the element, B_1;
- the element's initial expectation, $\mathrm{E}(B_1)$;
- the element's adjusted expectation for the data d actually seen, $\mathrm{E}_d(B_1)$;
- the element's standardised adjustment, $\mathrm{S}_d(B_1)$ (for an interpretation see section 5.3).

11.2 Consistency checks for the canonical directions

We can evaluate the adjusted expectation for each canonical direction and determine their corresponding standardised values. To see these we use the cda+ option:

BD>**show: cda+** ↩

giving the output shown in figure 17 (We discuss expectations for the canonical directions in section 5.4.) Each canonical direction has prior expectation zero and an adjusted expectation for the data d actually seen. Also shown is the magnitude of the difference in standard deviations. That is, recalling that each canonical direction has prior expectation zero, for each direction Y_i, we see $\mathrm{E}(Y_i) = 0$, $\mathrm{E}_d(Y_i)$, and $\mathrm{S}_d(Y_i)$.

Figure 17: Canonical direction expectations

Direction	Initial	Adjusted	Std
1	0.0000	0.4763	0.8440
2	0.0000	-0.3018	-2.1243

11.3 Viewing the size and direction of adjustment

We can examine the bearing and the standardised bearing for the current adjustment by using the b and b* arguments to the SHOW: command:

BD>**show: b** ↩
BD>**show: b*** ↩

Figure 18: The bearing

Element	Current
B1	0.3880
B2	0.1583
Constant	-2.6033
Variance	0.3179
Expected	0.3386
Ratio	0.9389

Figure 19: The standardised bearing

Element	Current
B1	0.4106
B2	0.2468
Variance	0.3179
Expected	0.3386
Ratio	0.9389

giving respectively the output shown in figure 18 and figure 19. This output (which we discuss further in section 5.5) consists of, for figure 18:

- the bearing $\mathrm{Z}_d(B)$, with the coefficients of B_1 and B_2 shown in addition to a constant;
- the size of the adjustment (equivalent to the bearing variance), $\mathrm{Size}_d(B)$;
- the expected size of the adjustment (which we can show to be equal to the resolution of the belief structure $[B]$, $\mathrm{R}_D(B)$);
- and the size ratio, $\mathrm{Sr}_D(B)$, being the ratio of the actual size of adjustment to the expected size.

and for figure 19:

- the same bearing given above, $\mathrm{Z}_d(B)$, expressed in terms of the quantities B_1 and B_2 standardised to have expectation zero and variance unity. (The constant term for the standardised bearing is zero, and is not output.)

12 Further [B/D] techniques

This section introduces [B/D] techniques that are a little more advanced than those that we have seen already. In particular we consider using [B/D] to transform some of our inputs and outputs in ways which allow a more detailed analysis of the example. If you wish only to generate the output referred to in part II, simply type the commands that are shown below in a bold typeface preceeded with the [B/D] prompt `BD>`, and ignore the commentary!

12.1 Generating the canonical directions

Prior to carrying out an adjustment, we can order [B/D] to retain[16] the canonical directions by issuing a KEEP: command with a cd argument. Usually we don't bother to keep the canonical directions, and retaining them unnecessarily slows the program a little, so to keep them now we will have to fix it so that the program knows that it is to retain the directions, and then re-enter the ADJUST: command:

`BD>`**keep:cd=Y** ↩
`BD>`**adjust:[B / D]** ↩

We have chosen the stem "Y" for our name, so that the names of our canonical directions will be Y1 and Y2. (We could have chosen another stem, such as "Mad", corresponding to naming the canonical directions as "Mad1" and "Mad2".) When the adjustment has been carried out, the directions will be retained and available to us in the form of **assignments**.

12.2 [B/D] Assignments

Assignments are, in [B/D], simple linear combinations of quantities. We can look at the assignments that we have just retained by entering the following:

`BD>`**look:(a)** ↩

Figure 20: Assignment definitions

```
Assignments :
Y2 =   0.3020 B1   - 0.7039 B2   + 3.1431
Y1 =   1.0059 B1   - 0.1136 B2   - 3.4745
```

giving the output shown in figure 20. What we are shown are all the assignments defined so far: these two directions. (Compare them with the canonical directions output in figure 14: they should be the same.) In future, we can continue to use the LOOK:(a) command to check on our assignments. The main use we make of assignments is in constructing further quantities, and we will illustrate how we do this by using the BUILD: command.

12.3 Building assignments

We have retained our canonical directions as assignments; as linear combinations. The point is that by retaining them in this way (and not as elements), we don't have to take up room by defining the covariance matrices that would associate them with the other quantities which are, as **elements**, automatically associated with specific variance-covariance matrices. Hence the assignment is a very efficient way of retaining some quantities.

However, sooner or later we may want to construct **elements** from the assignments that exist, and we do so using the BUILD: command. This command can be used in different ways, for different purposes. To construct our canonical directions, which exist as assignments Y1 and Y2, we use it as follows:

[16] Much of [B/D]'s output is, or can be, retained for further use, perhaps as fresh input to the program. Some quantities are retained automatically: these include summaries of the latest adjustment, for example. Other quantities can be retained if you set the relevant control.

BD>**build:Y1** $\hookleftarrow$
BD>**build:Y2** $\hookleftarrow$

Now we will use some of the commands that we used earlier in section 8 to examine our construction[17]. Issue the following command to inspect the names of elements and their associated covariance matrix:

BD>**look:(e,v1)** $\hookleftarrow$

Figure 21: Constructing the canonical directions

```
Element      Expectation
B1             4.160000
B2             6.250000
D1             4.160000
D2             6.250000
Y1             0.000000
Y2             0.000000
Covariances in store (1) :
B1 :     1.1200   0.7200   0.6200   0.3000   1.0448  -0.1686
B2 :     0.7200   2.4300   0.3000   0.4300   0.4482  -1.4930
D1 :     0.6200   0.3000   1.1200   0.7200   0.5896  -0.0239
D2 :     0.3000   0.4300   0.7200   2.4300   0.2529  -0.2121
Y1 :     1.0448   0.4482   0.5896   0.2529   1.0000  -0.0000
Y2 :    -0.1686  -1.4930  -0.0239  -0.2121  -0.0000   1.0000
```

giving the output shown in figure 21. There are several observations to be made about this sequence of commands and their resultant output:

- we combined the commands LOOK:(e) (for looking at elements) and LOOK:(v1) (for looking at beliefs) into the single command LOOK:(e,v1), which gave us both. If we like, we can always combine the arguments of the LOOK: command in this way: we use a comma to separate alternative arguments to the command.
- Y_1 and Y_2 are now known to [B/D] as elements called Y1 and Y2 respectively, and [B/D] has determined from their definition as linear combinations of pre-existent elements that each has expectation zero.
- we see the full covariance matrix specified over all six elements. [B/D] automatically deduced the necessary covariances over Y_1 and Y_2, and between them and the pre-existent elements, from the linear combinations supplied.
- we asserted above that a property of the canonical directions is that they are uncorrelated and have unit variances. These features are readily verifiable here.

12.4 Using the constructed quantities

As you might guess, once our constructs are defined we can use them just as though we had defined them from the keyboard. As an example, we will use them to emphasise some technical properties of the canonical directions and of the belief structure generated by the canonical directions. Our name for the particular belief structure $[Y] = [Y_1, Y_2]$ is the belief grid. Let us see what happens when we adjust the belief grid $[Y]$ by $[D]$, instead of adjusting the belief structure $[B]$ by $[D]$.

The following commands define a new collection, or base, Y to contain our belief grid, adjust the collection Y by the collection D, and then display various results of the adjustment:

BD>**base: Y = Y$** $\hookleftarrow$
BD>**keep: -cd** $\hookleftarrow$
BD>**adjust: [Y / D]** $\hookleftarrow$
BD>**show: v+, e*, a+** $\hookleftarrow$

giving the output shown in figure 22 (with further discussion in section 4.5). What should we notice about this sequence of commands and the output?

[17] The assignments Y1 and Y2 still exist, but we can forget them for the time being.

Figure 22: Adjusting the canonical directions

```
              -------- Variation -----------    Current
Element       Initial  Adjustment   Resolved   Resolution
Y1             1.0000    0.6816       0.3184     0.3184
Y2             1.0000    0.9798       0.0202     0.0202
[Y]            2         1.6614       0.3386     0.1693
Element          Adjusted Expectation (standardized)
Y1 :          0.6007 D1  - 0.0999 D2  + 0.0000
Y2 :          0.0454 D1  - 0.1559 D2  + 0.0000
               Expectations:
Element     Initial    Adjusted      Std
Y1          0.0000      0.4763      0.8440
Y2          0.0000     -0.3018     -2.1243
```

- we used our abbreviated syntax "Y$" to mean all those elements whose names start with the letter Y. At this point in time, [B/D] knows two such elements, $Y1$ and $Y2$, so the base Y is defined to be the collection $\{Y_1, Y_2\}$.
- we didn't want to retain the canonical directions for *this* adjustment, so we turned the option off. Otherwise our previous retentions would have been overwritten.
- we can combine output options for the SHOW: command by supplying several options to the command, separated by commas.
- the results of the adjustment over the belief grid are just as we asserted above, as they constitute fundamental properties of the belief grid. Each canonical direction has an initial variance of unity by default, and the same resolution seen in figure 14. Each has an evaluated adjusted expectation identical to that seen earlier, in figure 17.
- the results for the belief structure $[Y]$ are identical to those for the belief structure $[B]$. This is to be expected as the two are equivalent in the sense that the quantities $\{Y_1, Y_2\}$ that generate $[Y]$ also form a basis for $[B]$.
- we also see the adjusted expectation formula for each canonical direction.

12.5 Constructing the bearing

In section 12.1 we showed you how to retain the canonical directions as assignments and we went on to construct them as quantities. We can do the same with the bearing vector: it is, after all, a simple linear combination of elements in B. To retain the bearing we need to tell [B/D] before making the adjustment, and this is achieved by issuing the command

BD>**keep: b=Z** ↩

This line informs [B/D] that the bearing for a subsequent adjustment is to be retained as an assignment named "Z" (here, too, we can choose another name if we wish). Suppose that we now effect the creation and retention of the bearing as an assignment by issuing another adjustment command:

BD>**adjust: [B/D]** ↩

Firstly we can verify that [B/D] has added the newly created assignment that represents the bearing to the others that we created earlier, and saw in figure 20:
BD>**look: (a)** ↩
giving the output shown in figure 23. The first assignment that we see corresponds to the bearing that we saw for the first time in section 11.3, and we see also the assignments Y_1 and Y_2 retained earlier. Now, just as we did for the canonical directions, we will go on to build the bearing and then assess it as a quantity by displaying its variance and covariances with B_1 and B_2:

Figure 23: Assigning the bearing

```
Assignments :
Z  =    0.3880 B1   + 0.1583 B2   - 2.6033
Y2 =    0.3020 B1   - 0.7039 B2   + 3.1431
Y1 =    1.0059 B1   - 0.1136 B2   - 3.4745
```

BD>**build: Z** ↩
BD>**look: (v1) B, Z** ↩

Figure 24: Constructing the bearing

```
Covariances in store (1) :
B1 :     1.1200   0.7200   0.5485
B2 :     0.7200   2.4300   0.6640
Z  :     0.5485   0.6640   0.3179
```

giving the output shown in figure 24. This shows that we have indeed constructed Z, and have displayed the variance matrix (with a variance for Z and covariances between Z and B deduced by [B/D] from the linear combination supplied) for B_1, B_2 and Z. Notice how we modify the LOOK: command to restrict the output to the detail that we want: we add a list of bases and elements to the command[18].

We can use this output to examine the covariances between the bearing and the elements of B, and verify that these covariances represent the changes in expectation for each quantity due to the adjustment. We can also verify another property, namely that the size of the adjustment shown in figure 18 is equal to the prior variance of the linear combination representing the bearing.

13 Adjusting beliefs in stages

In [B/D] we adjust beliefs in stages by using the ADJUST: command to make an initial adjustment, and then to use a modified form of the ADJUST: command indicating to [B/D] that the following adjustment is to be made in addition to that which precedes it.[19]

13.1 The initial adjustment

Suppose that we adjust the belief structure $[B]$ by $[D_1]$ initially, and display some of the results available:

BD>**adjust: [B/D1]** ↩
BD>**show: cd+,v+,a+,b** ↩

giving the output shown in figure 25. (We discuss some of these results further in section 6.1.) We asked to see detailed output relating to the variances; evaluations of the adjusted expectations; the canonical directions and their resolutions; and the bearing, and the explanation of what the output represents is as we have discussed in earlier sections except, of course, that they refer to the simpler adjustment by $[D_1]$ only.

[18]The standard LOOK: command thus consists of a list of options in parentheses, optionally followed by a list of items, meaning that detail is to be restricted to these items. Such restrictions are relevant to only a subset of the possible options.

[19]There are other ways of asking [B/D] to perform sequences of adjustments, suited to different contexts. There are also facilities which help to automate the partial adjustment process for sequences in a manner similar to some traditional stepwise regression approaches.

Figure 25: Adjusting $[B]$ by $[D_1]$

```
  Canonical Directions :
                    1
Resolution       0.3109
B1               0.9928
B2              -0.0849
Constant        -3.5992
            -------- Variation -----------
Element     Initial  Adjustment   Resolved   Resolution
B1           1.1200    0.7768       0.3432     0.3064
B2           2.4300    2.3496       0.0804     0.0331
[B]          2         1.6891       0.3109     0.1554
             Expectations:
Element   Initial    Adjusted      Std
B1         4.1600     4.8464      1.1717
B2         6.2500     6.5821      1.1717
    Bearings :
Element              Current
B1                    0.6485
B2                   -0.0555
Constant             -2.3512
Variance              0.4268
Expected              0.3109
Ratio                 1.3729
```

13.2 Subsequent partial adjustments

Now suppose that we wish to adjust $[B]$ by $[D_2]$ in addition to $[D_1]$. We achieve this by issuing the following command:

BD>**adjust: [+/D2]** ↩

Here the "[+/D2]" syntax indicates that the previous adjustment is to be used as the basis for a new adjustment, and that $[D_2/D_1]$ (the portion of D_2 that is not common to D_1) is the additional source of information. Now we are ready to review some of the results available for such a partial adjustment.

13.2.1 Partial variances

To show the effects of partial adjustment upon the variances of the quantities of interest, we use the pv option to the SHOW: command:

BD>**show: pv** ↩

Figure 26: Partial adjustment variation

	-Partial adjustment variation-			Partial	Adj. Version
Element	Previous	Current	Resolved	Resolution	Resolution
B1	0.7768	0.7718	0.0049	0.0044	0.0064
B2	2.3496	2.3211	0.0286	0.0118	0.0122
[B]	1.6891	1.6614	0.0277	0.0139	0.0164

giving the output shown in figure 26. (This is discussed further in section 6.2.1.) Reading from left to right, and using B_1 to illustrate, we see

- the name of the element, B_1;
- the adjustment variance for the element after the initial adjustment, $\mathrm{Var}_{D_1}(B_1)$;

- the adjustment variance for the element taking into account the additional partial adjustment, that is $\text{Var}_{[D_1 \cup D_2]}(B_1) = \text{Var}_D(B_1)$;
- the difference between the first two quantities, representing the variance resolved by the partial adjustment, $\text{RVar}_{[D_2/D_1]}(B_1)$;
- the partial resolution, $\text{R}_{[D_2/D_1]}(B_1)$, being the proportion of variation resolved relative to the initial variance in B_1;
- the proportion of variance resolved by the partial adjustment, relative to the adjustment variance in B_1 prior to the partial adjustment. In terms of summarising the reduction in variance for $[B_1/D_1]$ (the adjusted version of B_1 given D_1), this is $\text{R}_{[D_2/D_1]}([B_1/D_1])$.

The final row of the output gives analagous results for uncertainties over the belief structure $[B]$ taken as a whole. Reading from left to right we see:

- the name of the structure, $[B]$;
- the adjustment uncertainty for the structure after the initial adjustment, $\text{U}_{D_1}(B)$;
- the adjustment uncertainty for the structure following the additional partial adjustment, that is $\text{U}_{[D_1 \cup D_2]}(B)$;
- the difference between the first two quantities, representing the uncertainty resolved by the partial adjustment, $\text{RU}_{[D_2/D_1]}(B)$;
- the partial resolution, $\text{R}_{[D_2/D_1]}(B)$;
- $\text{R}_{[D_2/D_1]}([B/D_1])$, where $[B/D_1]$ here is the belief structure $[B]$ adjusted by the belief structure $[D_1]$.

13.2.2 Partial canonical directions

Every partial adjustment implies consequences for the uncertainties of the quantities of interest, which we summarise in the same way that canonical directions summarise the consequences for the full adjustment. To see these partial canonical directions we use the `pcd+` option to the SHOW: command:

BD>**show: pcd+** ↩

Figure 27: Partial canonical directions

```
Partial  Canonical Directions :
                 1
Resolution     0.0277
B1             0.7971
B2            -0.6539
Constant       0.7710
```

giving the solitary direction W shown in figure 27. The number of such partial directions is the same as the additional number of quantities introduced for the partial adjustment. It is given together with its resolution, $\text{R}_{[D_2/D_1]}(W)$. (For further discussion see section 6.2.2.)

13.2.3 Further results for the partial adjustment

To display further details about the partial adjustment we can issue the usual SHOW: command with various arguments, including for the first time the `pc` option, which we use to display the path correlation:

BD>**show: cd+,v+,a+,b,pc** ↩

giving figure 28. The results displayed here are generally interpreted in the same way as before (in particular see sections 6.2.3, 6.2.4, and 6.2.5) but there are some extra details offered:

Figure 28: Adjusting $[B]$ additionally by $[D_2]$

```
  Canonical Directions :
                1          2
Resolution     0.3184     0.0202
B1             1.0059     0.3020
B2            -0.1136    -0.7039
Constant      -3.4745     3.1431
            -------- Variation -----------
Element     Initial  Adjustment   Resolved   Resolution
B1           1.1200    0.7718      0.3482     0.3109
B2           2.4300    2.3211      0.1089     0.0448
[B]          2         1.6614      0.3386     0.1693
             Expectations:
Element   Initial    Adjusted      Std     Change       Std
B1        4.1600     4.7085      0.9296   -0.1379     -1.9628
B2        6.2500     6.9140      2.0117    0.3319      1.9628
    Bearings :
Element              Current       Adjustment      Previous
B1                    0.3880       -0.2606          0.6485
B2                    0.1583        0.2138         -0.0555
Constant             -2.6033       -0.2521         -2.3512
Variance              0.3179        0.1069          0.4268
Expected              0.3386        0.0277          0.3109
Ratio                 0.9389        3.8524          1.3729
Path Correlation                              -0.5051
```

- the output displaying evaluations of the adjusted expectations contain two additional columns which show the actual change in expectation between the current and the previous adjustment; and this same change standardised relative to the partial variance resolved. Using the quantity B_1 to illustrate, these extra columns represent

$$\begin{aligned} \mathrm{E}_{[d_2/d_1]}(B_1) &= \mathrm{E}_d(B_1) - \mathrm{E}_{d_1}(B_1), \\ \text{and} \quad \mathrm{S}_{[d_2/d_1]}(B_1) &= \frac{\mathrm{E}_{[d_2/d_1]}(B_1)}{\mathrm{RVar}_{[D_2/D_1]}(B_1)}. \end{aligned}$$

- the bearing for the previous and current adjustments are given together with the bearing for the partial adjustment. Reading from left to right we have the coefficients for the linear combination representing the current bearing, $\mathrm{Z}_d(B)$, the bearing for the partial adjustment, $\mathrm{Z}_{[d_2/d_1]}(B)$, and the bearing for the initial adjustment by D_1, $\mathrm{Z}_{d_1}(B)$. You should be able to verify by inspection the property

$$\mathrm{Z}_d(B) = \mathrm{Z}_{d_1}(B) + \mathrm{Z}_{[d_2/d_1]}(B).$$

  Underneath this output can be found the size and expected size of each adjustment, together with the size ratio for each adjustment.

- the path correlation, $\mathrm{C}(d_1, [d_2/d_1])$, being the prior correlation between the initial and partial bearings.

13.3 Withdrawing quantities from the adjustment

We withdraw quantities from the adjustment in a way similar to that in which we added quantities. To remove the element D_1 from the adjustment, and to display various details about this withdrawal, issue the commands:

BD>**adjust: [-/D1]** ↩
BD>**show: cd+, v+, a+, b** ↩
BD>**show: pcd+, pv** ↩

Figure 29: Withdrawing D_1 from the adjustment

```
  Canonical Directions :
                 1
Resolution     0.0448
B1             0.5767
B2             0.3652
Constant      -4.6819
            -------- Variation -----------
Element     Initial  Adjustment   Resolved   Resolution
B1           1.1200    1.0830       0.0370     0.0331
B2           2.4300    2.3539       0.0761     0.0313
[B]          2         1.9552       0.0448     0.0224
             Expectations:
Element   Initial    Adjusted      Std     Change       Std
B1         4.1600     4.5983      2.2773    0.1102      0.1976
B2         6.2500     6.8782      2.2773    0.0358      0.1976
    Bearings :
Element             Current       Adjustment      Previous
B1                   0.2781        0.1099          0.3880
B2                   0.1761       -0.0178          0.1583
Constant            -2.2576       -0.3457         -2.6033
Variance             0.2325        0.0115          0.3179
Expected             0.0448        0.2938          0.3386
Ratio                5.1862        0.0390          0.9389
```

Figure 30: The partial adjustment $[D_1/D_2]$

```
Partial  Canonical Directions :
                 1
Resolution     0.2938
B1             1.0258
B2            -0.1663
Constant      -3.2277
            -Partial adjustment variation-    Partial     Adj. Version
Element      Previous  Current     Resolved   Resolution    Resolution
B1            0.7718   1.0830      -0.3111    -0.2778        -0.4031
B2            2.3211   2.3539      -0.0329    -0.0135        -0.0142
[B]           1.6614   1.9552      -0.2938    -0.1469        -0.1768
```

This time, the "[-/D1]" syntax indicates that the previous adjustment (which was over $[D]$ in full) is to be used as the basis for a new adjustment, and that $[D_1/D_2]$ (the portion of D_1 that is not common to D_2) is the source of information withdrawn.

The resulting output (which is considered further in section 6.3) should by now be self-explanatory. Figure 29 displays the canonical directions; detailed variance results; detailed expectation results; and the bearings for the simple adjustment by D_2 only that remains after withdrawing D_1. Additionally, there are some results pertaining to the change in adjustment - for example the bearing output shows the bearing for the partial adjustment.

Figure 30 shows the partial canonical direction and partial variances for the partial adjustment by $[D_1/D_2]$. Bear in mind here that the partial adjustment refers to what happens, and what is expected to happen, when a quantity is withdrawn from the adjustment. Thus, the variances for the quantities of interest are growing larger rather than smaller, and so forth. Hence, the partial variation results show negative resolutions representing increases in variance.

14 Review of [B/D] usage

To this point we have shown you how to start and stop [B/D]. We have used the following commands for basic organisation and input of a problem:

ELEMENT: to introduce uncertain quantities and their expectations;

BASE: to organise quantities into bases;

VAR: to specify beliefs in the form of variances and covariances for the uncertain quantities;

DATA: to specify data for the observable quantities.

To check upon our inputs using these commands, we used the LOOK: command with several different arguments:

e for elements;

b for bases;

v1 for variances;

r1 for correlations;

d for data.

When our specifications were completed, we issued the ADJUST: command, [B/D]'s most important command, which performs the general adjustment of one belief structure by another. The SHOW: command was then used to extract various information about the adjustment. In particular, the following display options were used:

- e and e* to see the adjusted expectations and standardised adjusted expectations;
- v and v+ to see adjusted variances for our quantities;
- a and a+ to see the evaluations of the adjusted expectations for the particular data that we observed;
- cd+ to see the canonical directions and their resolutions;
- cda+ to see evaluations of the adjusted expectations for the canonical directions;
- b and b* to see the bearing and standardised bearing, summarising the observed adjustment.

In addition, we introduced the notion of an **assignment** meaning a qauntity in [B/D] being represented as a linear combination. We saw how the canonical directions and the bearing can be retained as assignments by using the KEEP: command with arguments cd and b respectively; and we saw how they can be examined by using the LOOK: command with argument a.

We showed you how to use the BUILD: command to transform assignments into elements, and we constructed the canonical directions and the bearing from their representations as assignments into elements which became usable for further [B/D] analysis.

Finally, we considered some ways in which the [B/D] language is used to determine partial adjustments of belief, and we used the following options to see results specific to partial adjustments:

pv to see a summary of information about the effects partial adjustment on uncertainties.

pcd+ to see the partial canonical directions and their resolutions.

pc to display the path correlation.

Acknowledgements

All computations in this paper were carried out using the program [B/D], developed by us with the help of grants from the Science and Engineering Research Council. We are grateful to Malcolm Farrow for providing beliefs and data.

References

[1] M. Goldstein. Temporal coherence. In J.-M. Bernardo et al., editors, *Bayesian Statistics 2*, pages 231–248. North Holland, 1985.

[2] M. Goldstein. [B/D] introduction and overview. Technical Report 1:5, University of Hull Mathematics Research Reports, 1987.

[3] M. Goldstein. Can we build a subjectivist statistical package? In R. Viertl, editor, *Probability and Bayesian Statistics*, pages 203–217. Plenum, New York, 1987.

[4] M. Goldstein. Bayes linear methods I - Adjusting beliefs: concepts and properties. Technical Report 1995/1, Department of Mathematical Sciences, University of Durham, 1995.

[5] M. Goldstein and D. A. Wooff. Bayes linear computation: concepts, implementation and programming environment. *Statistics and Computing*, 5:327–341, 1995.

[6] D. A. Wooff. [B/D] reference manual. Technical Report 1:6, University of Hull Mathematics Research Reports, 1987.

[7] D. A. Wooff. [B/D] works. In J.-M. Bernardo et al., editors, *Bayesian Statistics 4*, pages 851–859. Oxford University Press, 1992.

[8] D. A. Wooff. [B/D] Manual. Technical Report 1995/4, Department of Mathematical Sciences, University of Durham, 1995.