Bayes Linear Methods I
Adjusting Beliefs: Concepts and Properties

Michael Goldstein
University of Durham

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

Abstract

Bayes linear methodology provides a quantitative structure for expressing our beliefs and systematic methods for revising these beliefs given observational data. Particular emphasis is placed upon interpretation of and diagnostics for the specification. The approach is similar in spirit to the standard Bayes analysis, but is constructed so as to avoid much of the burden of specification and computation of the full Bayes case. This report is the first of a series describing Bayes linear methods. In this document, we introduce some of the basic machinery of the theory. Examples, computational issues, detailed derivations of results and approaches to belief elicitation will be addressed in related reports.

1 Introduction

Bayes linear methodology provides a quantitative structure for expressing our beliefs and systematic methods for revising these beliefs given observational data. Particular emphasis is placed upon interpretive and diagnostic features of the analysis. The approach is similar in spirit to the standard Bayes analysis, but is constructed so as to avoid much of the burden of specification and computation of the full Bayes case. From a foundational view, the Bayes analysis emerges as a special case of the Bayes linear approach. From a practical view, Bayes linear methods offer a way of tackling problems which are too complex to be handled by standard Bayesian tools.

This report is the first of a series describing Bayes linear methods. In this document, we introduce some of the basic machinery of the theory. Examples, computational issues, detailed derivations of results and methods for belief elicitation will be addressed in related reports. In particular, [9] contains a simple tutorial guide to the material in this report, by means of a simple example, with details as to how the relevant calculations may be programmed in the computer language [B/D].

We cover the following material.

Section 2 concerns our basic approach to quantifying uncertainty and details the specification requirements for the Bayes linear analysis.

Section 3 defines and interprets the notions of *adjusted expectation* and *adjusted variance* for a collection of quantities, and explains the role of *canonical directions* in summarising the effects of an adjustment.

Section 4 concerns the types of diagnostic comparisons that we may make after we have evaluated the belief adjustment. In particular, we discuss the role of the *bearing* of the adjustment in summarising the overall magnitude and nature of the changes between prior and adjusted beliefs.

Section 5 covers the role of *partial adjustments* for analysing beliefs which are modified in stages.

Section 6 Bayes linear methods are so named as, formally, they derive their properties from the linear structure of inner product spaces rather than the boolean structure of probability spaces. This section summarises the geometry underlying the adjustment of beliefs.

2 Quantifying uncertainty

In a quantitative belief analysis, we quantify various aspects of our beliefs about a collection of unknown quantities, and then, typically, we use further information to modify our statements of belief about these quantities. In this section, we consider the structure within which we shall express initial uncertainties.

2.1 Quantifying uncertainty

There are many different ways in which beliefs may be quantified. Most familiar, perhaps, is the Bayesian approach, in which beliefs about all of the uncertain quantities of interest are represented in terms of a joint probability distribution. In practice, the specification of such a joint probability distribution will often be largely arbitrary due to the difficulty that most of us find in thinking meaningfully and consistently in high numbers of dimensions (or even in low numbers of dimensions - indeed even specifying a single probability may be a daunting task if our answer really matters for some purpose).

Full probabilistic specification is unwieldy as a fundamental expression of prior knowledge in that it requires such an extremely large number of statements of prior knowledge, expressing judgements to so fine a level of detail, that usually we have neither the interest nor the ability to make most of these judgements in a meaningful way. To escape from the straitjacket of full probabilistic specification, we suggest an approach which is related in spirit to the Bayesian approach, but is more straightforward to apply.

Suppose, therefore, that we intend to quantify some aspects of our prior judgements. It is reasonable to require that our subsequent analysis should only be based on those aspects of our beliefs which we are both willing and able to specify. Each number that we specify expresses some aspect of our prior knowledge, and as such requires careful consideration. Our concern is to develop a methodology which allows us to specify and analyse relatively small, carefully chosen collections of quantitative judgements about whichever aspects of a problem are within our ability to specify in a meaningful way.

We begin by describing our basic approach to the quantification of belief.

2.2 Expectation

When we reduce the number of aspects of uncertainty about which specifications are to be made, we may also simplify the nature of the specification process, by using methods which lead directly to the particular quantifications that we require. For this purpose, we make direct assessments for our (subjective) **expectations** for the various uncertainties of interest.

The idea of treating expectation as a primitive quantity and specifying expectation directly rather than through some intermediary probabilistic specification has been developed at length by various authors. The most detailed exposition of this approach is described in de Finetti ([1, 2]). De Finetti uses the term *prevision* for an expectation elicited directly and suggests various operational definitions for directly elicited expectations.[1] In this formulation, the probability of an event is simply the expectation or prevision for the associated indicator function.

We shall therefore assume, in what follows, that we have made various prior expectation statements, through direct elicitation. We cannot give formal rules for specifying prior expectations any more than we can give such rules for specifying prior probabilities in a standard Bayes analysis. Each expectation expresses a subjective choice that must be made given our assessment of the situation in question. Our account concerns the various methods by which we can improve our quantifications of belief, given such initial judgements and relevant data. Thus, while the forming of sensible prior judgements is of fundamental importance, it falls outside the strict remit of this account. We will discuss in a separate report the issues involved in eliciting such restricted prior specifications. All that we shall observe here is that, because any full probability specification over some outcome space is logically equivalent to a specification of the expectation for every random quantity which could possibly be constructed over that outcome space, it must be a substantially easier task to make a careful prior specification of the expectations only for a small subset of such quantities.

[1] For example, the simplest such definition is that your prevision for a random quantity X is the value x that you consider to be a "fair price" for a ticket which pays X.

2.3 Belief specification

In general, the level of detail at which we choose to describe our beliefs will depend on

- how interested we are in the various aspects of the problem;
- our ability to specify each aspect of our uncertainty;
- the amount of time and effort that we are willing to expend on the problem;
- how much detail is required from our prior specification in order to extract the necessary information from the data.

We must, therefore, recognise that our analysis depends not only upon the observed data but also upon the level of detail to which we have expressed our beliefs. The formal framework within which we shall express our judgements is as follows.

We begin by supplying an ordered (finite or infinite) list $C = \{X_1, X_2, \ldots\}$ of random quantities, for which we shall make statements of uncertainty. We call C the **base** for our analysis.

For each $X_i, X_j \in C$ we specify

1. the expectation, $\mathrm{E}(X_i)$, giving a simple quantification of our belief as to the magnitude of X_i;
2. the variance, $\mathrm{Var}(X_i)$, quantifying our uncertainty or degree of confidence in our judgements of the magnitude of X_i;
3. the covariance, $\mathrm{Cov}(X_i, X_j)$, expressing a judgement on the relationship between the quantities, quantifying the extent to which observation on X_j may (linearly) influence our belief as to the size of X_i.

These expectations, variances and covariances are specified directly, although this does not preclude us from deducing the values from some larger specification, or even, when this is practical, from a full prior probability distribution. We require that each element of C must have finite prior variance.

For any ordered subcollections, A, B, of elements of C, we write

$$\mathrm{Var}(A)$$

to denote the variance matrix of the vector of elements of A, and we write

$$\mathrm{Cov}(A, B)$$

to denote the covariance matrix between the vectors A and B.

We control the level of detail of our investigations by our choice of the collection C. The most detailed collection that we could possibly select would consist of the indicator functions for all of the combinations of possible values of all of the quantities of interest.With this choice of C, we obtain a full probability specification over some underlying outcome space. Sometimes this special case may be appropriate, but for large problems we will usually restrict attention to small subcollections of this collection. (Thus, for example, if there were two quantities Y and Z which we might measure, then C might contain the terms $\{Y, Y^2, Z, Z^2, YZ\}$.) It is preferable to work explicitly with the collection of belief specifications that we have actually made rather than to pretend to specify much larger collections of prior belief statements.

2.4 Belief structures

The formal structure which is described by our belief specification is as follows. We have a collection of random quantities $C = \{X_1, X_2, \ldots\}$, each with finite prior variance. We construct the linear space $\langle C \rangle$ consisting of all finite linear combinations

$$h_0 X_0 + h_1 X_{i_1} + \ldots + h_k X_{i_k}$$

of the elements of C, where X_0 is the unit constant. We view $\langle C \rangle$ as a vector space in which each X_i is a vector, and linear combinations of vectors are the corresponding linear combinations of the random quantities.

$\langle C\rangle$ is in general the largest structure over which expectations are defined once we have defined expectations for the elements of C.

Covariance defines an inner product $(\cdot,\cdot)$ and norm over $\langle C\rangle$, defined, for $X, Y \in \langle C\rangle$ to be

$$(X,Y) = \mathrm{Cov}(X,Y),\ \ \|X\|^2 = \mathrm{Var}(X).$$

The vector space, $\langle C\rangle$, with the covariance inner product $(.,.)$, defines an inner product space, which we denote $[C]$. We call $[C]$ a **belief structure with base** $\{C\}$.[2] In this space, the 'length' of any vector is equal to the standard deviation of the random quantity.

A belief structure provides the minimal formal structuring for a belief specification which is sufficient for our general analyses. A traditional discrete probability space is represented within this formulation by a base consisting of indicator functions over a partition, so that the vectors are the linear combinations of the indicator functions, or, equivalently, the random variables over the probability space. A continuous probability specification is expressed as the Hilbert space of square integrable functions over the space with respect to the prior measure. In the probability specification, all covariances between all such pairs of random quantities over the space must be specified. The belief structure allows us to restrict, by our choice of base, the specification to any linear subspace of this collection, so that we may specify only those aspects of our beliefs which we are both able and willing to quantify. Therefore, the formal properties of our approach follow from the linearity underlying the inner product structure, which is why we term our approach **Bayes linear.**

In the following sections, we describe various general properties of belief adjustment. In the final section, we return to the geometry underlying this approach, and describe the formal structure of the analysis.

3 Adjusting beliefs by data

In this section, we discuss the adjustment of a collection of expectation statements, given data. As this report is intended as a summary of concepts and properties, results will be stated without proof. Technical details will be discussed in a separate report. To simplify the exposition, we shall suppose that our chosen base C contains a finite number of quantities. In the final section, we will describe the underlying geometry, identify the equivalent results for infinite collections, and give geometric explanations for the various properties that we have described.

3.1 Adjusted expectation

We have a collection, C, of random quantities, for which we have specified prior means, variances and covariances. Suppose now that we observe the values of a subset, $D = \{D_1, \ldots, D_k\}$, of the members of C. We intend to modify our beliefs about the remaining quantities, $B = \{B_1, \ldots, B_r\}$, in C. A simple method by which we can modify our prior expectation statements is to evaluate the adjusted expectation for each quantity.

The **adjusted expectation** of a random quantity $X \in B$, given observation of a collection of quantities D, written $\mathrm{E}_D(X)$, is defined to be the linear combination

$$\mathrm{E}_D(X) = h_D^T D = \sum_{i=0}^{k} h_i D_i$$

which minimises

$$\mathrm{E}((X - \sum_{i=0}^{k} h_i D_i)^2),$$

over all collections $h = (h_0, h_1, \ldots, h_k)$, where $D_0 = 1$. $\mathrm{E}_D(X)$is sometimes called the **Bayes linear rule for X given D.**

[2]Strictly, the inner product space is defined over the closure of the equivalence classes of random quantities which differ by a constant, so that we identify any vector, such as X_0, with zero variance with the zero vector.

$\mathrm{E}_D(X)$ is determined by the prior mean, variance and covariance specifications. If $\mathrm{Var}(D)$ is of full rank[3] then

$$\mathrm{E}_D(X) = \mathrm{E}(X) + \mathrm{Cov}(X, D)[\mathrm{Var}(D)]^{-1}(D - \mathrm{E}(D)). \tag{1}$$

Adjusted expectation obeys the following properties:

1. for any quantities X_1, X_2 and constants c, d we have,

$$\mathrm{E}_D(cX_1 + dX_2) = c\mathrm{E}_D(X_1) + d\mathrm{E}_D(X_2) \tag{2}$$

2. for any X, we have

$$\mathrm{E}(\mathrm{E}_D(X)) = \mathrm{E}(X) \tag{3}$$

3.2 Interpretation

How should we interpret adjusted expectations? There are four inter-related interpretations that we can offer.

- The simplest interpretation is to view the quantity $\mathrm{E}_D(X)$ as an 'estimator' of the value of X, which combines the data with simple aspects of our prior beliefs in an intuitively plausible manner and which leads to a useful methodology. Alternately, if we have extensive data sources to draw upon, then we may construct our prior judgements from these sources and use our approach to develop 'estimators' which can be viewed as complementary to certain standard estimators in multivariate analysis.

- The second interpretation is to view adjusted expectation simply as a primitive which quantifies certain aspects of our beliefs, in a similar manner to the original expectation statement. Indeed, in de Finetti's formal development of prevision, the principle operational definition that he offers is that our prevision for X is the value x which we would choose if we were forced to suffer a penalty

$$L = k(X - x)^2, \tag{4}$$

where k is a constant defining the units of loss. In this view adjusted expectation simply expresses the extension of our choice of preferences from the certain choice x to the random choice

$$L_D = k(X - \sum_{i=0}^{k} x_i D_i)^2. \tag{5}$$

In the special case where the elements D_i are the indicator functions for a partition, then this is equivalent to de Finetti's choice for the operational definition of the conditional prevision, x_i, of X given each event D_i. Under this view, adjusted expectations are simply informative summaries, generalising the corresponding conditional expectations defined over indicator functions.

- If we are committed in principle to a full Bayes view based on complete probabilistic specification of all uncertainties, then we may view adjusted expectations as offering simple tractable approximations to a full Bayes analysis for complicated problems. In addition, the various interpretive measures and diagnostic tests which we shall develop below offer insights which are relevant to any full Bayes analysis.

- We have described three alternative views of adjusted expectation, each of which has merit in certain contexts and reflects the contrasting views that may be held concerning the revision of beliefs. However, we hold a fourth view, which, by proceeding directly by foundational arguments, subsumes each of the above views. This view explains why we should view adjusted expectation as a primitive, the precise sense in which adjusted expectation may be viewed as an 'estimator', and the general properties which may be claimed for the estimate. Further, it reverses our third interpretation above by identifying a full Bayes analysis as a simple special case of the general analysis which we advocate.

[3] If $\mathrm{Var}(D)$ is not of full rank, then we may discard elements of D so that the reduced collection is of full rank. Otherwise, we may consider $[\mathrm{Var}(D)]^{-1}$ to be the Moore-Penrose generalised inverse in the following matrix equations.

Our immediate intention is to describe the practical machinery of our approach. Therefore, we do not at this point intend to take logical and philosophical diversions into foundational issues, and we shall develop the formal relationship between belief adjustment and belief revision elsewhere. Instead, for now we will move between the first three interpretations that we have listed above, viewing adjusted expectation as an intuitively plausible numerical summary statement about our beliefs given the data. There is no implication that this value will fully express our genuine revised belief concerning the expectation of X. Rather, we have been explicit as to precisely which aspects of our prior beliefs have been utilised in order to assess the adjusted expectation. As with any other formal analysis that we might carry out, adjusted expectations offer logical information in quantitative form which we may use as we deem appropriate to improve our actual posterior judgements.

3.3 Adjusted variance

We define the **adjusted version** of X given D, $[X/D]$, to be the 'residual' form

$$[X/D] = X - \mathrm{E}_D(X).$$

Adjusted quantities obey the following properties:

1. $$\mathrm{E}([X/D]) = 0; \tag{6}$$

2. $$\mathrm{Cov}([X/D], \mathrm{E}_D(X)) = 0. \tag{7}$$

We write X as the sum of the two uncorrelated components

$$X = [X/D] + \mathrm{E}_D(X),$$

so that we can split $\mathrm{Var}(X)$ as

$$\mathrm{Var}(X) = \mathrm{Var}([X/D]) + \mathrm{Var}(\mathrm{E}_D(X)).$$

The **variance of the adjusted version of** X, or the **adjusted variance**, $\mathrm{Var}_D(X)$, is defined to be

$$\mathrm{Var}_D(X) = \mathrm{Var}([X/D]) = \mathrm{E}((X - \mathrm{E}_D(X))^2).$$

The value of $\mathrm{Var}_D(X)$ is determined by our prior variances and covariances as

$$\mathrm{Var}_D(X) = \mathrm{Var}(X) - \mathrm{Cov}(X, D)[\mathrm{Var}(D)]^{-1}\mathrm{Cov}(D, X) \tag{8}$$

The **variance of X resolved by** D, $\mathrm{RVar}_D(X)$, is defined as

$$\mathrm{RVar}_D(X) = \mathrm{Var}(\mathrm{E}_D(X)) = \mathrm{Cov}(X, D)[\mathrm{Var}(D)]^{-1}\mathrm{Cov}(D, X).$$

We therefore write the variance partition for X as

$$\mathrm{Var}(X) = \mathrm{Var}_D(X) + \mathrm{RVar}_D(X) \tag{9}$$

In line with our various interpretations of belief adjustment, we may give corresponding interpretations to adjusted variance. We may view $\mathrm{Var}_D(X)$ as:

- the 'mean squared error' of the estimator $\mathrm{E}_D(X)$;
- a primitive expression, interpreted as we would a prior variance, but applied to the 'residual variation' when we have extracted the variation in X 'accounted for' by D;
- a simple, easily computable upper bound on full Bayes preposterior risk, under quadratic loss, for any full prior specification consistent with the given mean and variance specifications;

- within the more general view of the foundations, the adjustment variance attaches directly to our posterior beliefs.

We quantify the effect of an adjustment by evaluating the **resolution**, $\mathrm{R}_D(X)$, defined as

$$\mathrm{R}_D(X) = \frac{\mathrm{RVar}_D(X)}{\mathrm{Var}(X)} = 1 - \frac{\mathrm{Var}_D(X)}{\mathrm{Var}(X)}. \tag{10}$$

If $\mathrm{R}_D(X)$ is near zero then either the collection D is not expected to be informative for X, relative to our prior knowledge about X, or our beliefs have not been specified in sufficient detail to exploit the information contained in D.

Finally, we define the **adjusted covariance**, $\mathrm{Cov}_D(X,Y)$ to be

$$\mathrm{Cov}_D(X,Y) = \mathrm{Cov}([X/D],[Y/D]) = \mathrm{E}((X - \mathrm{E}_D(X))(Y - \mathrm{E}_D(Y))).$$

3.4 Adjusting a collection of quantities

We have suggested how we might adjust our prior expectation for any one element of a collection $B = \{B_1, \dots, B_r\}$ using observations on a collection $D = \{D_1, \dots, D_k\}$. When we evaluate a collection of adjusted expectations $\{\mathrm{E}_D(B_1), \dots, \mathrm{E}_D(B_k)\}$, we also implicitly evaluate the adjusted value for each element of $\langle B\rangle$, the collection of linear combinations of the elements of B, as, by the linearity of adjusted expectation (equation 2),

$$\mathrm{E}_D(\sum_{i=1}^{r} h_i B_i) = \sum_{i=1}^{r} h_i \mathrm{E}_D(B_i). \tag{11}$$

We now analyse changes in beliefs over $\langle B\rangle$. We consider B, D as vectors, of dimension r and k, respectively. We define the adjusted version of the collection B given D, $[B/D]$, to be the 'residual' vector

$$[B/D] = B - \mathrm{E}_D(B).$$

The properties of the adjusted vector are as for a single quantity, namely

1. $$\mathrm{E}([B/D]) = 0, \tag{12}$$
   the r dimensional null vector,
2. $$\mathrm{Cov}(\mathrm{E}_D(B),[B/D]) = 0, \tag{13}$$
   the $r \times k$ null matrix.

Therefore, just as for a single quantity X, we partition the vector B as the sum of two uncorrelated vectors, namely

$$B = \mathrm{E}_D(B) + [B/D], \tag{14}$$

so that we may partition the variance matrix of B into two variance components

$$\mathrm{Var}(B) = \mathrm{Var}(\mathrm{E}_D(B)) + \mathrm{Var}([B/D]) \tag{15}$$

We call

$$\mathrm{RVar}_D(B) = \mathrm{Var}(\mathrm{E}_D(B)),$$

the **resolved variance matrix, for** B **by** D. We call

$$\mathrm{Var}_D(B) = \mathrm{Var}([B/D])$$

the **adjusted variance matrix, for** B **by** D.

$\mathrm{E}_D(B)$, $\mathrm{Var}_D(B)$ are calculated as in equations 1, 8, namely

$$\mathrm{E}_D(B) = \mathrm{E}(B) + \mathrm{Cov}(B,D)[\mathrm{Var}(D)]^{-1}(D - \mathrm{E}(D)), \tag{16}$$
$$\mathrm{Var}_D(B) = \mathrm{Var}(B) - \mathrm{Cov}(B,D)[\mathrm{Var}(D)]^{-1}\mathrm{Cov}(D,B), \tag{17}$$
$$\mathrm{RVar}_D(B) = \mathrm{Cov}(B,D)[\mathrm{Var}(D)]^{-1}\mathrm{Cov}(D,B). \tag{18}$$

3.5 Adjusted belief structures

If we adjust each member of the base $\{B\}$ by D, then we obtain a new base $\{[B_1/D], \ldots, [B_k/D]\}$, the base of adjusted versions of the elements of B. We call this the **base** $\{B\}$ **adjusted by** D, written $\mathbf{\{B/D\}}$. The belief structure with this base is termed the **adjusted belief structure of** B **by** D and is written $[B/D]$. To simplify our notation, we also use $[B/D]$ to represent the vector $([B_1/D], \ldots, [B_k/D])$, where appropriate.

We may view $[B/D]$ as representing a belief structure over the linear space $\langle\{B/D\}\rangle$. However, it is also useful to view $[B/D]$ as an inner product space constructed over the linear space $\langle B\rangle$ but with the covariance inner product replaced by the adjusted covariance inner product

$$(X, Y)_D = \mathrm{Cov}_D(X, Y) = \mathrm{Cov}([X/D], [Y/D]).$$

We now analyse the differences between the variance and the adjusted variance inner products.

3.6 Canonical directions

To assess how much information about $[B]$ we expect to receive by observing D, we may first identify the particular linear combination $Y_1 \in \langle B\rangle$ for which we expect the adjustment by D to be most informative in the sense that Y_1 maximises the resolution $\mathrm{R}_D(Y)$ over all elements $Y \in \langle B\rangle$ with non-zero prior variance. (Note from equation 10, that maximising the resolution is equivalent to minimising the ratio of adjusted to prior variance.) We may then proceed to identify directions for which we expect progressively less information. This is equivalent to defining collections of canonical variables between $[B]$ and $[D]$. We make the following definition.

DEFINITION The j^{th} **canonical direction for the adjustment of** B **by** D is the linear combination Y_j which maximises $\mathrm{R}_D(Y)$ over all elements $Y \in \langle B\rangle$ with non-zero prior variance which are uncorrelated a priori with $Y_1, \ldots, Y_{j-1}$. We scale each Y_j to have prior expectation zero and prior variance one. The values

$$r_i = \mathrm{R}_D(Y_i) = \mathrm{RVar}_D(Y_i)$$

are termed the **canonical resolutions**. The number of canonical directions that we may define is equal to the rank, $r(B)$, of the variance matrix of the elements of B.

The quantities $\{Y_1, \ldots, Y_{r(B)}\}$ are mutually incorrelated. It is also the case that the adjusted expectations, $\{\mathrm{E}_D(Y_1), \ldots, \mathrm{E}_D(Y_{r(B)})\}$ are also mutually uncorrelated, and each Y_i is uncorrelated with each $\mathrm{E}_D(Y_j), j \neq i$.

The canonical resolutions may be calculated as the eigenvalues of the **resolution matrix**, T_D, defined as

$$T_D = [\mathrm{Var}(B)]^{-1}\mathrm{RVar}_D(B). \tag{19}$$

We may calculate $Y_1, \ldots, Y_{r(B)}$ by finding the normed eigenvectors, $v_1, \ldots, v_{r(B)}$, ordered by eigenvalues $1 \geq r_1 \geq r_2 \geq \ldots \geq r_{r(B)} \geq 0$, of T_D, so that

$$Y_i = v_i^T B, \ \mathrm{Var}_D(Y_i) = 1 - r_i.$$

The collection $\{Y_1, Y_2, \ldots\}$ forms a "grid" of directions over $\langle B\rangle$, summarising the effects of the adjustment. We expect to learn most about those linear combinations of the elements of B which have large correlations with those canonical directions with large resolutions. The exact relation is as follows.

For any X in $\langle B\rangle$,

$$\mathrm{R}_D(X) = \sum_{i=1}^{r(B)} c_i(X) r_i, \tag{20}$$

where

$$c_i(X) = \frac{[Corr(X, Y_i)]^2}{\sum_{j=1}^{r(B)}[Corr(X, Y_j)]^2}.$$

3.7 The system resolution uncertainty

By analogy with the resolution for a single random quantity, we define the **resolved uncertainty for the belief structure** [**B**] to be

$$\mathrm{RU}_D(B) = \sum_{i=1}^{r(B)} r_i.$$

The total resolution is the sum of the resolutions for any collection of $r(B)$ elements of $\langle B \rangle$ with prior variance one, which are a priori uncorrelated. Note that if D is the empty set $\emptyset$, then

$$\mathrm{RU}_\emptyset(B) = r(B).$$

Where appropriate, we may therefore view $r(B)$ as the **prior uncertainty in the collection** B, written as

$$\mathrm{U}(B) = r(B),$$

namely the total uncertainty associated with any maximal collection of uncorrelated elements of $\langle B \rangle$ standardised to unit variance. We define the **system resolution for** B to be the ratio of resolved to prior uncertainty for the collection, namely

$$\mathrm{R}_D(B) = \frac{\mathrm{RU}_D(B)}{\mathrm{U}(B)}.$$

The system resolution provides qualitatively similar information for the structure $[B]$ to that expressed by the resolution for a single quantity, X. $\mathrm{R}_D(B)$ can be viewed as the "average" of the resolutions for each canonical direction, so that a value near one implies that we expect substantial information about most elements of $\langle B \rangle$, while a value near zero indicates that there are a variety of elements for which the adjustment is not expected to be informative.

4 The observed adjustment

In the previous section, we constructed adjusted expectations given collections of observations. After we make the observations and evaluate these adjustments, we express our overall changes in belief in ways which help us both to identify qualitatively the most important changes between our prior and adjusted beliefs, and also to judge diagnostically whether we should re-examine any aspects of our formulation. We proceed as follows.

4.1 Standardised observations

Each prior statement that we make describes our beliefs about some random quantity. When we actually observe this quantity, we may compare what we expect to happen with what actually happens. A simple comparison is as follows.

For a single random quantity X, suppose that we specify $\mathrm{E}(X)$ and $\mathrm{Var}(X)$ and then observe value x. Using only our limited belief specification, we evaluate the standardised observation defined as

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

$\mathrm{S}(x)$ has prior expectation zero and prior variance one. Thus, a very large absolute value for $\mathrm{S}(x)$ might suggest that we have misspecified $\mathrm{E}(X)$ or underestimated the variability of X, or misrecorded the value x, while a value near zero might suggest that we have overestimated the variability of X. How large or small $\mathrm{S}(x)$ must be to merit attention depends entirely upon the context, relating in large part to our confidence in our prior formulation.

4.2 Standardised adjustments

Suppose that we specify beliefs about a quantity, X, then adjust these beliefs by observation on a collection of quantities, D. When we observe the actual data values,

$$D = d = (d_1, \ldots, d_k),$$

then we may evaluate the random quantity $\mathrm{E}_D(X)$. The value which is obtained is denoted by $\mathrm{E}_d(X)$. We apply the standardisation operation to $\mathrm{E}_d(X)$, defining the **standardised adjustment** as

$$\mathrm{S}_d(X) = \mathrm{S}(\mathrm{E}_d(X)) = \frac{\mathrm{E}_d(X) - \mathrm{E}(X)}{\sqrt{\mathrm{RVar}_D(X)}}.$$

The value of $\mathrm{S}_d(X)$ may suggest that our beliefs about X appear to be more or less affected by the data than we had expected. Very large changes may raise the possibility that we have been overly confident in describing our uncertainty, very small changes that we have been overly modest in valuing our prior knowledge about the value of X.

Such diagnostics provide us with qualitative and quantitative information. If our observations suggest to us substantially new beliefs, then presumably it will be of interest to us to know this. (For example, we may appear to have made a great discovery simply because of a blunder in our programming). Even when no simple explanation of a possible discrepancy occurs to us, it will usually be of interest to identify which aspects of our beliefs have changed by substantially less or more than we had expected. Such diagnostics are of particular importance when we make very large collections of belief adjustments, so that we need simple, automatic methods to call our attention to particular assessments which we might usefully re-examine.

4.3 Canonical standardised adjustments

When we adjust a collection, B, of random quantities by a further collection D, there are many standardised adjustments that we may evaluate. A systematic collection of such consistency checks on our specification is provided by evaluating the standardised value for each of the canonical directions, Y_i, for the adjustment. We term these values the **canonical standardised adjustments** defined as

$$\mathrm{S}_d(Y_i) = \frac{\mathrm{E}_d(Y_i) - \mathrm{E}(Y_i)}{\sqrt{r_i}}. \tag{21}$$

There are two types of diagnostic information given by these values. Quantitatively, any aberrant value may require scrutiny. Qualitatively, we may look for systematic patterns. For example, as we expect larger changes in belief for the first canonical directions than for subsequent directions, a particularly revealing pattern would be a sequence of decreasing absolute values, which might suggest qualitatively a false prior classification between the more and the less informative directions.

4.4 The bearing of the adjustment

Each evaluation that we have so far discussed assesses the change in belief for a single element of $\langle B \rangle$. We now summarise our overall changes in belief over $\langle B \rangle$, relative to our prior uncertainty. We make the following definition.

Definition The **size** of the adjustment of B given $D = d$ is

$$\mathrm{Size}_d(B) = \max_{X \in \langle B \rangle} \frac{(\mathrm{E}_d(X) - \mathrm{E}(X))^2}{\mathrm{Var}(X)}.$$

Note: There are various alternative scalings for the changes in belief which we can choose, each of which may be analysed in a similar fashion to our suggested choice and provide useful insights into the belief revision. Our particular choice leads to the construction of various quantities whose properties unify many of the interpretive and diagnostic features of the belief revision, and is particularly helpful when we come to consider the adjustment of beliefs in stages.

We now identify the element, $\mathrm{Z}_d(B)$, of $\langle B\rangle$ with the largest such change in expectation. $\mathrm{Z}_d(B)$ is termed the **bearing** for the adjustment, and is constructed as follows.

Definition The **bearing** for the adjustment of the belief structure $[B]$ by observation of $D = d$ is the element $\mathrm{Z}_d(B)$ in $\langle B\rangle$ defined by

$$\mathrm{Z}_d(B) = \sum_{i=1}^{r(B)} \mathrm{E}_d(U_i)U_i, \tag{22}$$

where $U_1, \ldots, U_{r(B)}$ are any collection of elements of $\langle B\rangle$ which are a priori uncorrelated, with variance one. (The canonical components of $\mathrm{Var}(B)$ form one such collection and the canonical directions for the adjustment form another when suitably scaled. $\mathrm{Z}_d(B)$ does not depend on the choice of $U_1, \ldots, U_{r(B)}$.)

The bearing is so named as it expresses both the direction and the magnitude of the change between prior and adjusted beliefs, as follows:

1. for any X which is a priori uncorrelated with $\mathrm{Z}_d(B)$, $\mathrm{E}_d(X) = \mathrm{E}(X)$;
2. if $M_d = \alpha \mathrm{Z}_d(B)$, then a bearing of M_d would represent α times the change in expectation as would a bearing of $\mathrm{Z}_d(B)$, for every element of $\langle B\rangle$.
3. these properties follow as

$$\mathrm{E}_d(X) - \mathrm{E}(X) = \mathrm{Cov}(X, \mathrm{Z}_d(B)), \ \forall X \in \langle B\rangle. \tag{23}$$

We may therefore deduce that $\mathrm{Z}_d(B)$ is indeed the direction of maximum change in belief, and that

$$\mathrm{Size}_d(B) = \mathrm{Var}(\mathrm{Z}_d(B)) = \sum_{i=1}^{r(B)} E_d^2(U_i). \tag{24}$$

4.5 The expected size of an adjustment

A natural diagnostic for assessing the magnitude of an adjustment is to compare the largest standardised change in expectation that we observe to our expectation for the magnitude of the largest change, evaluated prior to observing D. This expectation is assessed as follows.

$$\mathrm{E}(\mathrm{Size}_D(B)) = \mathrm{E}(\mathrm{Var}(\mathrm{Z}_D(B))) = \mathrm{RU}_D(B), \tag{25}$$

($\mathrm{Z}_D(B)$ is the random element of $\langle B\rangle$ which takes the value $\mathrm{Z}_d(B)$ if D takes value d.) Thus, the expected size of the adjustment is equal to the resolved uncertainty for the structure. To compare the observed and expected values, we define the **size ratio for the adjustment of B by D** to be

$$\mathrm{Sr}_d(B) = \frac{\mathrm{Size}_d(B)}{\mathrm{E}(\mathrm{Size}_D(B))} = \frac{\mathrm{Var}(\mathrm{Z}_d(B))}{\mathrm{RU}_D(B)}. \tag{26}$$

We anticipate that the ratio will be near one. Large values of the size ratio suggest that we have formed new beliefs which are surprisingly discordant with our prior judgements. Values near zero might suggest that we have exaggerated our prior uncertainty.

The size ratio is essentially a ratio of variances. To determine some 'critical size' for this quantity, we would, at the least, need to assess the variance of our variance statements, i.e. to make fourth moment specifications. For the present, we treat the ratio as a simple warning flag drawing our attention to possible conflicts between prior and adjusted beliefs.[4]

[4] As an example of the type of simple rule of thumb that might sometimes be of use, observe that were all the elements of D to be normally distributed, then it would follow that

$$\mathrm{Var}(\mathrm{Sr}_D(B)) = 2\frac{\sum_{i=1}^{r(B)} r_i^2}{(\sum_{i=1}^{r(B)} r_i)^2}.$$

In certain circumstances, we might find it useful to approximate the distribution of $\mathrm{Sr}_D(B)$, for example by a distribution of form cX, where c is a constant and X has a χ^2 distribution, with ν degrees of freedom. Matching the mean and variance suggests a choice of $\nu = (\sum_{i=1}^{r(B)} r_i)^2/(\sum_{i=1}^{r(B)} r_i^2)$ degrees of freedom and $c = 1/\nu$.

4.6 Data size

Earlier in this section, we considered standardised changes in various individual quantities. We then considered measures of maximal discrepancy in adjusted expectation. We now combine these two assessments.

For any data vector D, we may construct the collection of linear combinations $\langle D\rangle$. For any element $F \in \langle D\rangle$, with observed value f, we must have $\mathrm{E}_d(F) = f$. Therefore the element of $\langle D\rangle$ with the largest standardised observation

$$\max_{F\in\langle D\rangle} (\frac{f - \mathrm{E}(F)}{\sqrt{\mathrm{Var}(F)}})^2,$$

is precisely the bearing $\mathrm{Z}_d(D)$ of the adjustment of D by D. We therefore define the size of the data observation D as follows.

Definition The **size** of the data vector $D = d$ is

$$\mathrm{Size}_d(D) = \max_{F\in\langle D\rangle} (\mathrm{S}(f))^2.$$

$\mathrm{Size}_d(D)$ is as defined in subsection 4.4. As in that subsection, we may construct the quantity $\mathrm{Z}_d(D)$ as

$$\mathrm{Z}_d(D) = \sum_{i=1}^{r(D)} u_i U_i, \tag{27}$$

where $U_1, \ldots, U_{r(D)}$ are any uncorrelated collection of elements of $\langle D\rangle$, with prior variance one, and u_i is the observed value of U_i. $\mathrm{Z}_d(D)$ has the property that for any $F \in \langle D\rangle$

$$f - \mathrm{E}(F) = \mathrm{Cov}(F, \mathrm{Z}_d(D)). \tag{28}$$

We have

$$\mathrm{Size}_d(D) = \mathrm{Var}(\mathrm{Z}_d(D)) = \sum_{i=1}^{r(D)} u_i^2,$$

and

$$\mathrm{E}(\mathrm{Size}_D(D)) = r(D),$$

the rank of $\mathrm{Var}(D)$. The size ratio for data vector $D = d$ is therefore

$$\mathrm{Sr}_d(D) = \frac{\mathrm{Size}_d(D)}{\mathrm{E}(\mathrm{Size}_D(D))} = \frac{\sum_{i=1}^{r(D)} u_i^2}{r(D)}.$$

Again, we expect this value to be near one. Values which are very large or very close to zero suggest similar possible misspecifications to those for a general adjustment.

5 Adjusting beliefs in stages

We have described the adjustment of beliefs about a collection of quantities by observation of a further collection. Often, we will want to explore the ways in which different aspects of the data and the prior specification combine to give the final adjustment. (For example, we might be combining information of various different types collected in different places by different people at different times.) We now consider which aspects of the data are most crucial to the final adjustment, in order to produce efficient sampling frames and experimental designs, a priori, and to investigate diagnostically whether the various portions of the observed data have similar or contradictory effects on our beliefs.

5.1 Partial adjustment of beliefs

Suppose that we intend to adjust our beliefs about a collection $B = \{B_1, \ldots, B_r\}$ by observation of two further collections $D = \{D_1, \ldots, D_k\}$ and $F = \{F_1, \ldots, F_j\}$ of quantities. We adjust B by the collection $(D \cup F)$ (i.e. the collection $\{D_1, \ldots, D_k, F_1, \ldots, F_j\}$) but separate the effects of the subsets of data. Therefore, we adjust B in stages, first by D, then adding F. We may show that the additional adjustment of B by F, given that we have already adjusted by D, is the same as the adjustment of B by $[F/D]$, the belief structure $[F]$ adjusted by $[D]$:

$$\mathrm{E}_{(D \cup F)}(B) - \mathrm{E}_D(B) = \mathrm{E}_{[F/D]}(B), \tag{29}$$

(This relation follows as adjusting F by D removes the 'common variability' between F and D.) We call

$$\mathrm{E}_{[F/D]}(B),$$

the **(partial) adjustment of B by F given D**. Note the following properties of partial adjustment.

1.
$$\mathrm{E}(\mathrm{E}_{[F/D]}(B)) = 0. \tag{30}$$

2.
$$\mathrm{E}_{D \cup F}(B) = \mathrm{E}_D(B) + \mathrm{E}_{[F/D]}(B). \tag{31}$$

3.
$$[B/(D \cup F)] = [(B/D)/(F/D)]. \tag{32}$$

4.
$$\begin{aligned} \mathrm{Cov}(\mathrm{E}_{[F/D]}(B), \mathrm{E}_D(B)) &= \mathrm{Cov}(\mathrm{E}_{[F/D]}(B), [B/(F \cup D)]) \\ &= \mathrm{Cov}([B/(F \cup D)], \mathrm{E}_D(B)) = 0. \end{aligned} \tag{33}$$

5.2 Partial variance

In section 3, we split the vector B into two uncorrelated components, as

$$B = \mathrm{E}_D(B) + (B - \mathrm{E}_D(B)).$$

We further decompose $(B - \mathrm{E}_D(B))$, and write

$$\begin{aligned} B &= \mathrm{E}_D(B) + (\mathrm{E}_{D \cup F}(B) - \mathrm{E}_D(B)) + (B - \mathrm{E}_{D \cup F}(B)) \\ &= \mathrm{E}_D(B) + \mathrm{E}_{[F/D]}(B) + [B/F \cup D]. \end{aligned} \tag{34}$$

The three components on the right hand side of the above equation are mutually uncorrelated. We may partition $\mathrm{Var}_D(B)$, the 'unresolved variation' from the adjustment by D, as

$$\mathrm{Var}_D(B) = \mathrm{Var}(\mathrm{E}_{[F/D]}(B)) + \mathrm{Var}_{(D \cup F)}(B). \tag{35}$$

The second term is the adjusted variance matrix of B by $D \cup F$, and the first is the **(partial) resolved variance matrix of B by F given D,** namely

$$\mathrm{RVar}_{[F/D]}(B) = \mathrm{Var}(\mathrm{E}_{[F/D]}(B)).$$

Resolved variances are additive in the sense that

$$\mathrm{RVar}_{(F \cup D)}(B) = \mathrm{RVar}_D(B) + \mathrm{RVar}_{[F/D]}(B).$$

For any $X \in \langle B \rangle$, we assess the further reduction in 'residual variation' from adding F, given D, as the **(partial) resolution**, namely

$$\mathrm{R}_{[F/D]}(X) = \frac{\mathrm{RVar}_{[F/D]}(X)}{\mathrm{Var}(X)}. \tag{36}$$

5.3 Partial canonical directions

We summarise the effects of the partial adjustment in a similar fashion to that for a full adjustment. We make the following definition.

DEFINITION The j^{th} **partial canonical direction for the adjustment of** B **by** F **given** D is the linear combination W_j which maximises $\mathrm{R}_{[F/D]}(B)$ over all elements in $\langle B\rangle$ with non-zero prior variance which are uncorrelated with each $W_i,\ i<j$, scaled so that each $\mathrm{Var}(W_j)=1$. The values

$$f_i = R_{[F/D]}(W_i)$$

are termed the **partial canonical resolutions**.

The partial canonical directions for F given D are evaluated exactly as are the canonical directions for D, as described in subsection 3.6, but the eigenstructure is extracted from the **partial resolution matrix**

$$T_{[F/D]} = [\mathrm{Var}(B)]^{-1}\mathrm{RVar}_{[F/D]}(B).$$

The collection $W_1, W_2\ldots$ forms a "grid" of directions over $\langle B\rangle$, summarising the additional effects of the adjustment. Having adjusted by D, we expect to learn most additionally from F for those linear combinations of the elements of B which have large correlations with those partial canonical directions with large resolutions. The exact relation is as before, namely for any $X \in \langle B\rangle$,

$$\mathrm{R}_{[F/D]}(X) = \sum_{i=1}^{r(B)} c_i(X) f_i, \tag{37}$$

where

$$c_i(X) = \frac{(Corr(X, W_i))^2}{\sum_{j=1}^{r(B)} (Corr(X, W_j))^2}.$$

The **system partial resolution** is

$$\mathrm{R}_{[F/D]}(B) = \frac{\sum_{i=1}^{r(B)} f_i}{r(B)}$$

The resolution is additive, namely

$$\mathrm{R}_D(B) + \mathrm{R}_{[F/D]}(B) = \mathrm{R}_{D\cup F}(B).$$

When we have made the adjustment, in addition to evaluating canonical standardised adjustments for the adjustment by D and by $D \cup F$, we may obtain similar qualitative insights into the changes in adjustment by evaluating the **partial canonical standardised adjustments** which are as in subsection 4.3 but applied to the adjustment by $[F/D]$.

5.4 Representing the observed partial adjustment

When we observe the values of D and F, and so of $[F/D]$, taking values

$$D = d,\ F = f,\ [F/D] = [f/d],$$

then we may evaluate the **size of the partial adjustment** defined to be

$$\mathrm{Size}_{[f/d]}(B) = \max_{X \in \langle B\rangle} \frac{(\mathrm{E}_{d\cup f}(X) - \mathrm{E}_d(X))^2}{\mathrm{Var}(X)} = \max_{X \in \langle B\rangle} \frac{(\mathrm{E}_{[f/d]}(X))^2}{\mathrm{Var}(X)}.$$

We create the **bearing for the partial adjustment**, as

$$\mathrm{Z}_{[f/d]}(B) = \sum_{i=1}^{r(B)} \mathrm{E}_{[f/d]}(U_i) U_i,$$

for any collection $U_1, \ldots, U_{r(B)}$ mutually uncorrelated with unit prior variance. $\mathrm{Z}_{[f/d]}(B)$ satisfies the relation

$$\mathrm{E}_{d\cup f}(X) - \mathrm{E}_d(X) = \mathrm{E}_{[f/d]}(X) = \mathrm{Cov}(X, \mathrm{Z}_{[f/d]}(B)),\ \forall X \in \langle B\rangle. \tag{38}$$

Therefore

$$\mathrm{Size}_{[f/d]}(B) = \mathrm{Var}(\mathrm{Z}_{[f/d]}(B)),$$

which may be compared to the expected value, namely

$$\mathrm{E}(\mathrm{Size}_{[F/D]}(B)) = \mathrm{RU}_{[F/D]}(B).$$

Replacing B by F in the above expressions allows us to define the corresponding partial data size of F given D, namely the largest change

$$\mathrm{Size}_{[f/d]}(D) = \max_{F\in\langle F\rangle} \frac{(f - \mathrm{E}_D(F))^2}{\mathrm{Var}(F)} = \mathrm{Var}(\mathrm{Z}_{[f/d]}(F)).$$

5.5 Path correlation

When we adjust beliefs in stages, the expected sizes of the respective adjustments are additive in the sense that

$$\mathrm{E}(\mathrm{Size}_{D\cup F}(B)) = \mathrm{E}(\mathrm{Size}_D(B)) + \mathrm{E}(\mathrm{Size}_{[F/D]}(B))$$

However, the observed sizes of the adjustments are not additive. We have

$$\mathrm{Z}_{d\cup f}(B) = \mathrm{Z}_d(B) + \mathrm{Z}_{[f/d]}(B). \tag{39}$$

The size of each adjustment is the variance of the corresponding bearing. Therefore

$$\mathrm{Var}(\mathrm{Z}_{d\cup f}(B)) = \mathrm{Var}(\mathrm{Z}_d(B)) + \mathrm{Var}(\mathrm{Z}_{[f/d]}(B)) + 2\mathrm{Cov}(\mathrm{Z}_d(B), \mathrm{Z}_{[f/d]}(B)) \tag{40}$$

so that

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

Thus, while

$$\mathrm{E}(\mathrm{Cov}(\mathrm{Z}_D(B), \mathrm{Z}_{[F/D]}(B))) = 0,$$

the observed value of this covariance

$$\mathrm{Cov}(\mathrm{Z}_d(B), \mathrm{Z}_{[f/d]}(B))$$

may be taken to expresses the degree of support/conflict between the two collections of evidence in determining the revision of beliefs. As a summary, we define the **path correlation** to be

$$\mathrm{C}(d, [f/d]) = \mathrm{Corr}(\mathrm{Z}_d(B), \mathrm{Z}_{[f/d]}(B))$$

If this correlation is near 1 then the two collections of data are complementary in that their combined effect in changing our beliefs is greater than the sum of the individual effect of each collection. If the path correlation is near -1 then the two collections are giving contradictory messages which give smaller overall changes in belief, in combination, than we would expect from the individual adjustments with D and $[F/D]$.

5.6 Adjustment in several stages

Now suppose that we intend to adjust B sequentially by the collections of quantities $G_1, G_2, \ldots, G_m$. We define the cumulative collection

$$G_{[i]} = \bigcup_{j=1}^{i} G_j.$$

and denote the cumulative adjustment

$$\mathrm{E}_{[i]}(B) = \mathrm{E}_{G_{[i]}}(B).$$

We may 'partial out' any stage of the adjustment, defining for any $i > j$ the **partial adjustment of B by $G_{[i]}$ given $G_{[j]}$** as

$$\mathrm{E}_{[i/j]}(B) = \mathrm{E}_{[i]}(B) - \mathrm{E}_{[j]}(B) = \mathrm{E}_{[\bigcup_{k=j+1}^{i} G_k / \bigcup_{k=1}^{j} G_k]}(B)$$

Corresponding to the adjustment $\mathrm{E}_{[i]}(B)$ is the bearing $\mathrm{Z}_{[i]}(B)$. The bearing for the partial adjustment $\mathrm{E}_{[i/j]}(B)$ is

$$\mathrm{Z}_{[i/j]}(B) = \mathrm{Z}_{[i]}(B) - \mathrm{Z}_{[j]}(B).$$

As before, we have

$$\mathrm{E}_{[i]}(X) - \mathrm{E}_{[j]}(X) = \mathrm{Cov}(X, \mathrm{Z}_{[i/j]}(B)).$$

The bearing for the partial adjustment expresses the change in both magnitude and direction in beliefs between stages $[j]$ and $[i]$. The i^{th} **stepwise partial adjustment,** $\mathrm{E}_{[i/]}(B)$ is

$$\mathrm{E}_{[i/]}(B) = \mathrm{E}_{[i/i-1]}(B) = \mathrm{E}_{[G_i/G_{[i-1]}]}(B),$$

with bearing $\mathrm{Z}_{[i/]}(B) = \mathrm{Z}_{[i]}(B) - \mathrm{Z}_{[i-1]}(B)$. We refer to the full sequence of stepwise adjusted bearings $\mathrm{Z}_{[1]}(B), \mathrm{Z}_{[2/]}(B), \ldots, \mathrm{Z}_{[m/]}(B)$ as the **data trajectory**. For each j we may write

$$\mathrm{Z}_{[j]}(B) = \mathrm{Z}_{[1]}(B) + \mathrm{Z}_{[2/]}(B) + \ldots + \mathrm{Z}_{[j/]}(B)$$

We have therefore have that

$$\mathrm{Size}_{[j]}(B) = \mathrm{Size}_{[1]}(B) + \mathrm{Size}_{[2/]}(B) + \ldots + \mathrm{Size}_{[m/]}(B) + 2(C_{[2]} + \ldots + C_{[j]})$$

where

$$C_{[r]} = \mathrm{Cov}(\mathrm{Z}_{[r-1]}(B), \mathrm{Z}_{[r/]}(B))$$

So to examine the ways in which the individual terms combine to determine the revision we must consider

- the prior expectation for each change to assess which subcollections of data are expected to be informative;
- the individual adjusted bearings $\mathrm{Z}_{[i/]}(B)$ to identify the stages at which larger than expected changes in belief occur;
- the path correlations $C_{[i]}$ to see whether the evidence is internally supportive or contradictory.

6 The geometry of belief adjustment

Just as traditional Bayes methods derive their formal properties from the structure of probability spaces, Bayes linear methods derive their formal properties from the linear structure of inner product spaces. We now describe this underlying geometry.

6.1 Belief adjustment

We have defined a (partial) belief structure as follows:

We have a collection $C = \{X_1, X_2, \ldots\}$, finite or infinite, of random quantities, each with finite prior variance. We construct the vector space $\langle C \rangle$ consisting of all finite linear combinations

$$c_0 X_0 + c_1 X_{i_1} + \ldots + c_k X_{i_k}$$

of the elements of C, where X_0 is the unit constant. Covariance defines an inner product $(\cdot,\cdot)$ and norm, over the closure of the equivalence classes of random quantities which differ by a constant in $\langle C \rangle$, defined, for $X, Y \in \langle C \rangle$ to be

$$(X, Y) = \mathrm{Cov}(X, Y), \ \ \|X\|^2 = \mathrm{Var}(X).$$

The space $\langle C \rangle$ with covariance inner product is denoted as $[C]$, the (partial) belief structure with base $\{C\}$.

Belief adjustment is represented within this structure as follows:

We have a collection $\{C\} = \{B_1, B_2, \ldots, D_1, D_2, \ldots\}$, the base for our analysis. We construct $[C]$ as above. We construct the two subspaces $[B]$ and $[D]$ corresponding to bases $\{B\} = \{B_1, B_2, \ldots\}$ and $\{D\} = \{D_1, D_2, \ldots\}$. We define P_D to be the orthogonal projection from $[B]$ to $[D]$. Thus, for any $X \in \langle B \rangle$, $\mathrm{P}_D(X)$ is the element of $[D]$ which is closest to X in the variance norm. This orthogonal projection is therefore equivalent to the adjusted expectation, i.e.

$$\mathrm{E}_D(X) = \mathrm{P}_D(X). \tag{41}$$

Thus the adjusted version of X is

$$[X/D] = X - \mathrm{P}_D(X),$$

namely the perpendicular vector from X to the subspace $[D]$. The adjustment variance $\mathrm{Var}_D(X)$ is therefore equal to the squared perpendicular distance from X to $[D]$. Further, as

$$X = [X/D] + P_D(X)$$

and $[X/D]$ is perpendicular to $\mathrm{P}_D(X)$, we have

$$\|X\|^2 = \|[X/D]\|^2 + \|P_D(X)\|^2$$

which is the variance partition expressed in equation 9.

If we adjust each member of $\{B\}$ by D, we obtain a new base $\{[B_1/D], \ldots, [B_k/D]\}$, which we write as $\{B/D\}$. We use $[B/D]$ to represent both the vector of elements of $\{B/D\}$ and the adjusted belief structure of B by D.

Alternately, it is often useful to identify $[B/D]$ as a subspace of the overall inner product space $[B \cup D]$, namely the orthogonal complement of $[D]$ in $[B \cup D]$.

Note from this latter representation that for any bases D and F we may write a direct sum decomposition of $[D \cup F]$ into orthogonal subspaces as

$$[D \cup F] = [D] \oplus [F/D], \tag{42}$$

Therefore, we may write

$$\mathrm{P}_{[D \cup F]}(X) = \mathrm{P}_{[D]}(X) + \mathrm{P}_{[F/D]}(X), \ \ \forall X \in \langle B \rangle, \tag{43}$$

where the two projections on the right hand side of equation 43 are mutually orthogonal. The variance partition for a partial belief adjustment follows directly from this representation.

6.2 A comment on the choice of inner product

While it is natural to view the variance inner product as describing our uncertainties, we may choose any inner product over $\langle C \rangle$ which describes relevant aspects of our beliefs as the starting point for our analysis. One particular choice that is frequently useful is the **product inner product**,

$$(X, Y) = E(XY).$$

This inner product does not set the unit constant X_0 to zero. We can represent our original expectations by means of orthogonal projections onto the subspace generated by the unit constant as

$$\mathrm{E}(X) = \mathrm{P}_{X_0}(X). \tag{44}$$

Within this belief structure, the covariance inner product is simply the adjustment by the unit constant, so that the inner product space that we have termed $[B]$ above, under this representation is more fully expressed as $[B/X_0]$. Equivalently $[B]$ is the orthogonal complement of X_0 in $\langle B \rangle$ under the product inner product. Usually, we suppress the prior adjustment by X_0 for notational simplicity, illustrating our freedom to choose whichever inner product is appropriate to emphasise the important features of a particular analysis.

6.3 Belief transforms

Geometrically, the effect of the belief adjustment may be represented by the eigenstructure of a certain linear operator T_D defined on $[B]$. This operator T_D is defined to be

$$T_D = P_B P_D \tag{45}$$

where P_B, P_D are the orthogonal projections from $[D]$ to $[B]$, and from $[B]$ to $[D]$, respectively.

T_D is a bounded self-adjoint operator, as P_B, P_D are adjoint transforms, namely

$$(X, P_B(Y)) = (P_D(X), Y), \ \forall X \in [B], \ Y \in [D], \tag{46}$$

because both sides of the above equation are equal to (X,Y).

The operator T_D is termed the **resolution transform** for B induced by D, as it represents the variance resolved for each X by D as

$$\mathrm{RVar}_D(X) = \mathrm{Cov}(X, T_D(X)), \tag{47}$$

as

$$\mathrm{RVar}_D(X) = \mathrm{Var}(P_D(X)) = (P_D(X), P_D(X)) = (X, P_B(P_D(X))).$$

We may also evaluate the transform

$$S_D = I - T_D,$$

where I is the identity operator on $[B]$. We term S_D the **variance transform** for B induced by D, as adjusted covariance is represented by the relation, for each X and Y in $\langle B \rangle$, that

$$\mathrm{Cov}_D(X, Y) = \mathrm{Cov}(X, S_D Y), \tag{48}$$

or equivalently, in terms of the inner products over $[B]$, as

$$(X, Y)_D = (X, S_D Y). \tag{49}$$

T_D, S_D are self-adjoint operators, of norm at most one. They have common eigenvectors, Y_i, with eigenvalues $1 \geq r_i, s_i \geq 0$, where $r_i + s_i = 1$.

From equation 47, we may deduce that, provided T_D has a discrete spectrum, each canonical direction, Y_i, of the adjustment of B by D, is an eigenvector of T_D, with eigenvalue r_i, and conversely each eigenvector of T_D is a canonical direction of the adjustment. Thus the eigenstructure of T_D summarises the effects of the adjustment over the whole structure $[B]$. In particular, the resolved uncertainty may be written as

$$\mathrm{RU}_D(B) = \mathrm{Trace}(T_D). \tag{50}$$

6.4 Comparing inner products

The variance transform and the resolution transform are particular examples of the general class of **belief transforms**. Suppose that we specify two inner products $\{\cdot,\cdot\}_1$,$\{\cdot,\cdot\}_2$, over $\langle B\rangle$, derived perhaps from alternative prior formulations or alternative sampling frames. Provided that

$$\sup_{X\in\langle B\rangle}\frac{\{X,X\}_2}{\{X,X\}_1}=M_{12}<\infty, \tag{51}$$

then we may define a bounded, self-adjoint transform T on $\langle B\rangle$, under inner product $\{\cdot,\cdot\}_1$, with norm M_{12}, for which

$$\{X,Y\}_2=\{X,T(Y)\}_1,\ \forall X,Y\in\langle B\rangle. \tag{52}$$

T is termed the belief transform for $\{\cdot,\cdot\}_1$, associated with $\{\cdot,\cdot\}_2$. For example, the variance transform S_D is obtained by selecting $\{\cdot,\cdot\}_1$ to be the inner product $\mathrm{Cov}(X,Y)$, and $\{\cdot,\cdot\}_2$ to be the adjusted covariance inner product $\mathrm{Cov}_D(X,Y)$, so that

$$\mathrm{Cov}_D(X,Y)=\mathrm{Cov}(X,S_D(Y)),\ \forall X,Y\in\langle B\rangle.$$

Just as the eigenstructure of the variance transform summarises the comparison between the prior and adjusted variance specification, so does the eigenstructure of a general belief transform summarise the comparison between any two inner products. The ratio $\{X,X\}_2/\{X,X\}_1$ will be large/ near one / small according as whether X has large components corresponding to eigenvectors with large/ near one / small eigenvalues.

Belief transforms provide a natural way to compare sequences of inner products, as they are multiplicative. Let T_{ij} be the belief transform for $\{\cdot,\cdot\}_i$ associated with $\{\cdot,\cdot\}_j$. Then we have

$$T_{13}=T_{12}T_{23}, \tag{53}$$

(operator multiplication is by composition, namely $T_{12}T_{23}(X)=T_{12}(T_{23}(X))$), as

$$\{X,T_{13}(Y)\}_1=\{X,Y\}_3=\{X,T_{23}(Y)\}_2=\{X,T_{12}(T_{23}(Y))\}_1.$$

This relation allows us to decompose a particular comparison into constituent stages. For example, if we wish to adjust $[B]$ by $[D\cup F]$, then we may decompose the overall variance transform $S_{[D\cup F]}$, into the product

$$S_{[D\cup F]}=S_DS_{(D)F}, \tag{54}$$

where $S_{(D)F}$ is the variance transform S_F applied to the adjusted space $[B/D]$, so that

$$\mathrm{Cov}_{D\cup F}(X,Y)=\mathrm{Cov}_D(X,S_{(D)F}(Y)). \tag{55}$$

Such multiplicative forms offer a natural sequential construction for a complicated belief transform. They also allow us to apply the collection of interpretive and diagnostic tools that we have developed to each stage of a belief comparison or adjustment.

6.5 The bearing

By the Riesz representation for linear functionals, f is a bounded linear functional on $[B]$ if and only if there is a unique element $Z_f\in[B]$, for which

$$f(X)=(X,Z_f),\ \forall X\in\langle B\rangle.$$

The difference between the prior expectation $\mathrm{E}(X)$ and the observed adjusted expectation $\mathrm{E}_d(X)$ defines a linear functional

$$f_d(X)=\mathrm{E}_d(X)-\mathrm{E}(X),$$

on $[B]$. Therefore by the Riesz representation, if $f_d(X)$ is bounded on $[B]$[5], then there is a unique element $Z_d \in [B]$, corresponding to $f_d(X)$, for which

$$\mathrm{E}_d(X) - \mathrm{E}(X) = f_d(X) = (X, Z_d) = \mathrm{Cov}(X, Z_d).$$

This element is precisely the bearing as created in section 4, and the properties of the bearing may be deduced directly from this representation. Note that in the preceding sections we have also used the Riesz representation to create the bearing for two other functionals, namely the difference functional, $\mathrm{E}_{d\cup F}(X) - \mathrm{E}_d(X)$, and also the functional which replaces each X by its observed value.

References

[1] B. de Finetti. *Theory of probability, vol. 1.* Wiley, New York, 1974.

[2] B. de Finetti. *Theory of probability, vol. 2.* Wiley, New York, 1975.

[3] M. Goldstein. Revising previsions: a geometric interpretation. *J. R. Statist. Soc.*, B:43:105–130, 1981.

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

[5] M. Goldstein. Separating beliefs. In P.K. Goel and A. Zellner, editors, *Bayesian Inference and Decision Techniques.* North Holland, Amsterdam, 1986.

[6] M. Goldstein. Adjusting belief structures. *J. R. Statist. Soc.*, B:50:133–154, 1988.

[7] M. Goldstein. The data trajectory. In J.-M. Bernardo et al., editors, *Bayesian Statistics 3*, pages 189–209. Oxford University Press, 1988.

[8] M. Goldstein. Belief transforms and the comparison of hypotheses. *The Annals of Statistics*, 19:2067–2089, 1991.

[9] D. A. Wooff. Bayes linear methods II - An example with an introduction to [B/D]. Technical Report 1995/2, Department of Mathematical Sciences, University of Durham, 1995.

[5]for example, f_d will automatically be bounded if D has a finite number of elements