Augmented Bayes linear/Bayes graphical models

Simon C Shaw & Michael Goldstein
Department of Mathematical Sciences
University of Durham, UK

March 26, 2002

ABSTRACT

We describe a method for Bayesian augmentation of Bayes linear models, introducing a graphical model, termed a Bayes linear/Bayes (BLB) graphical model, where the data nodes are all fully specified, whilst the parameters of the model are only partially specified. The connections of the data to the parameters are also fully specified. The propagation of a single data observation is considered and revised expectations and variances obtained for the parameters using a mixture of conditioning and Bayes linear belief adjustment. Propagation of data from multiple nodes is then developed and a local computation algorithm, developing upon that of Goldstein & Wilkinson (2000), for the global revision of the BLB model given. The method is illustrated by an example of partition testing in software.

1 Introduction - a simple example

Consider the following example. We have two random quantities of interest, $\mathcal{B}_1$ and $\mathcal{B}_2$. $\mathcal{B}_1$ has two possible outcomes, -1 and 1 which we judge to be equally likely to occur. $\mathcal{B}_2$ is continuous on the interval $[0,1]$ and we assert that $E(\mathcal{B}_2) = 1/2$ and $Var(\mathcal{B}_2) = 1/20$. A judgement is made that $Cov(\mathcal{B}_1, \mathcal{B}_2) = 1/18$. The full probability specification for $\mathcal{B}_1$ may be viewed as equivalent to specifying the second-order structure of the collection of random quantities $\mathcal{B}_{1,P} = \{\mathcal{E}_{\mathcal{B}_1,1},\ \mathcal{E}_{\mathcal{B}_1,2}\}$, where $\mathcal{E}_{\mathcal{B}_1,1}$ is the event that $\mathcal{B}_1 = -1$ and $\mathcal{E}_{\mathcal{B}_1,2}$ the event that $\mathcal{B}_1 = 1$. Notice that the given covariance between $\mathcal{B}_1$ and $\mathcal{B}_2$ does not allow us to deduce the covariance between $\mathcal{E}_{\mathcal{B}_1,1}$ and $\mathcal{B}_2$ nor that between $\mathcal{E}_{\mathcal{B}_1,2}$ and $\mathcal{B}_2$.

Suppose that a third random quantity, $\mathcal{D}_1$, is to be observed and following its observation we would like to revise our beliefs over the collection $\mathcal{B} = \{\mathcal{B}_1,\ \mathcal{B}_2\}$. $\mathcal{D}_1$ can take the possible values -1, 0 and 1 and we are willing to make the following conditional probability specification for $\mathcal{D}_1$:

$$P(\mathcal{D}_1 = -1|\mathcal{B}_1 = -1) = 0.5;\ P(\mathcal{D}_1 = 0|\mathcal{B}_1 = -1) = 0.5;\ P(\mathcal{D}_1 = 1|\mathcal{B}_1 = -1) = 0; \quad (1)$$

$$P(\mathcal{D}_1 = -1|\mathcal{B}_1 = 1) = 0;\ P(\mathcal{D}_1 = 0|\mathcal{B}_1 = 1) = 0.5;\ P(\mathcal{D}_1 = 1|\mathcal{B}_1 = 1) = 0.5. \quad (2)$$

Thus, the full joint probability distribution of $\mathcal{B}_1$, $\mathcal{D}_1$ may be calculated. The marginal specification for $\mathcal{D}_1$ may be viewed as the specification of the second-order structure of the collection $\mathcal{D}_{1,P} = \{\mathcal{E}_{\mathcal{D}_1,1},\ \mathcal{E}_{\mathcal{D}_1,2},\ \mathcal{E}_{\mathcal{D}_1,3}\}$, where $\mathcal{E}_{\mathcal{D}_1,1}$ is the event that $\mathcal{D}_1 = -1$, $\mathcal{E}_{\mathcal{D}_1,2}$ the event that $\mathcal{D}_1 = 0$ and $\mathcal{E}_{\mathcal{D}_1,3}$ the event that $\mathcal{D}_1 = 1$. We are only willing to make a partial specification between $\mathcal{D}_1$ and $\mathcal{B}_2$, declaring that $Cov(\mathcal{D}_{1,P}, \mathcal{B}_2) = (-1/72\ 0\ 1/72)^T$.

How might we revise our beliefs over $\mathcal{B}$ following the observation of $\mathcal{D}_1$? Note that since only the full joint probability specification between $\mathcal{B}_1$ and $\mathcal{D}_1$ is specified, a full posterior analysis of $\mathcal{B}$, via Bayes theorem, is not possible. There is, however, a full second-order specification between $\mathcal{B}$ and $\mathcal{D}_{1,P}$ and so a Bayes linear analysis, see Goldstein (1999) for an overview of the approach, is possible. The advantage of this approach is that we may restrict our attention to those random quantities we are explicitly interested in revising in the light of new information, in this example the collection $\mathcal{B}$ following the receipt of $\mathcal{D}_1$, and a full specification is not required. This approach, however, does not fully utilise the information available to us in our prior specification. In particular, our adjusted variances over $\mathcal{B}$ do not depend upon the actual value of $\mathcal{D}_1$ observed. The conditional specification of $\mathcal{D}_1$ given $\mathcal{B}_1$, see equations (1) and (2), show that if $\mathcal{D}_1 = \pm 1$ then $\mathcal{B}_1 = \mathcal{D}_1$ with certainty and we'd like to revise the variance of $\mathcal{B}_1$ down to zero to reflect this. In terms of $\mathcal{B}_2$, it is as if we observed both $\mathcal{D}_1$ and $\mathcal{B}_1$. If $\mathcal{D}_1 = 0$, then since $Cov(\mathcal{E}_{\mathcal{D}_{1,2}}, \mathcal{B}_2) = 0$, we'd expect this observation to have no impact upon our beliefs about $\mathcal{B}_2$. Similarly, the symmetry of equations (1) and (2) lead us to a similar impression about $\mathcal{B}_1$. Is there a systematic way we can capture this intuition without making any further specifications? Can we make use of the doctrine of Bayes linear methods whilst allowing modifications to the adjustments to better utilise the prior specification?

2 The geometric interpretation of expectation

Consider a general collection of random quantities, $\mathcal{B} = \{\mathcal{B}_1, \mathcal{B}_2, \ldots\}$, of interest. We treat expectation as primitive and for each $\mathcal{B}_i \in \mathcal{B}$ we specify directly the prior expectation, $E(\mathcal{B}_i)$. We follow the development of de Finetti (1974, 1975) and construct the linear space $\langle \mathcal{B} \rangle$ consisting of all finite linear combinations of the elements of $\mathcal{B}$ with the unit constant, $\mathcal{B}_0$, added. A typical element of $\langle \mathcal{B} \rangle$ is thus $X = b_0 + \sum_u b_u \mathcal{B}_{v_u}$, where $\{v_1, v_2, \ldots\}$ is a general finite set of integers and $b_0 = b_0 \mathcal{B}_0$.

We view $\langle \mathcal{B} \rangle$ as a vector space by considering each $\mathcal{B}_i$ as a vector and linear combinations, X, of random quantities as the corresponding linear combinations of vectors. Thus, $\langle \mathcal{B} \rangle$ is the space of all quantities whose expectation is uniquely determined from the specification of expectations for all of the random quantities contained in $\mathcal{B}$. Having fixed the linear structure, we add the geometric framework by forming the inner product space $[\mathcal{B}]$ from the minimal closure of $\langle \mathcal{B} \rangle$ by imposing the following inner product and norm for X, $Y \in \langle \mathcal{B} \rangle$

$$(X, Y) = Cov(X, Y); \tag{3}$$
$$\|X\|^2 = Var(X). \tag{4}$$

The inner product space is defined over the closure of the equivalence classes of random quantities which differ by a constant. For example, the quantities X and $X^* = X - E(X)$ are equivalent in this representation. We may then follow the convention of standardising each random quantity by subtracting its prior expectation. We also restrict attention to random quantities, X, with $E(X^2) < \infty$. The inner product space $[\mathcal{B}]$ is termed a belief structure; see Goldstein (1986a). Two subspaces $[\mathcal{B}^*]$ and $[\mathcal{B}^\dagger]$, are orthogonal, written $[\mathcal{B}^*] \perp [\mathcal{B}^\dagger]$, if every element of the collection $\mathcal{B}^*$ is uncorrelated with every element of the collection $\mathcal{B}^\dagger$.

The belief structure construction allows you to restrict specification to whatever subspace of the full probabilistic structure you deem relevant. For example, it may contain only those random quantities that you are interested in revising beliefs about in the light of new information. A discrete probability space over a collection of random quantities, $\mathcal{B}$, may be represented by imposing the inner product and norm, equations (3) and (4), over the collection, $\mathcal{B}_P$, of indicator functions for the elementary events of $\mathcal{B}$. A continuous probability specification is represented by letting $\mathcal{B}_P$ be the collection of all functions of $\mathcal{B}$ which are square-integrable with respect to the prior measure f over $\mathcal{B}$ and then constructing the inner product over the linear space $\langle \mathcal{B}_P \rangle$.

If $\mathcal{B}$ is finite, say $\mathcal{B} = \{\mathcal{B}_1, \ldots, \mathcal{B}_r\}$, then we consider $\mathcal{B}$ to be the $r \times 1$ vector $[\mathcal{B}_1 \ldots \mathcal{B}_r]^T$. Our expectation statements are collected together as the $r \times 1$ vector $E(\mathcal{B})$, and the inner product space $[\mathcal{B}]$ is represented by the $r \times r$ matrix $Var(\mathcal{B})$. For typical elements $X = b_0 + b^T\mathcal{B}$, $Y = \tilde{b}_0 + \tilde{b}^T\mathcal{B}$, we have $E(X) = b_0 + b^T E(\mathcal{B})$ and $Cov(X, Y) = b^T Var(\mathcal{B})\tilde{b}$.

3 Bayes linear methods

Suppose that we are to receive the values of a data collection, $\mathcal{D} = \{\mathcal{D}_1, \mathcal{D}_2, \ldots\}$ and the receipt of this information will cause us to revise the values of the expectations and covariances we had assigned for the elements in $\langle \mathcal{B} \rangle$.

The most familiar revision of beliefs is the full posterior analysis using conditioning and Bayes theorem. For the collection of random quantities $\mathcal{B}$, and observed data $\mathcal{D}$ we construct the covariance inner product, or belief structure, over the linear space $\langle \mathcal{B}_P \cup \mathcal{D}_P \rangle$ and denote it by $[\mathcal{B}_P \cup \mathcal{D}_P]$. For any X, $Y \in \langle \mathcal{B}_P \rangle$, $E(X)$ is revised to the conditional expectation, $E(X|\mathcal{D})$, and, $Cov(X, Y)$ is revised to the conditional covariance, $Cov(X, Y|\mathcal{D})$. We may summarise this adjustment as the replacement of the covariance inner product, $[\mathcal{B}_P]$ with the conditional covariance, which we denote by $[\mathcal{B}_P|\mathcal{D}]$.

An alternative approach, involving the analysis of limited aspects of the full probability structure, is to calculate our adjusted beliefs using Bayes linear methods. For any X, $Y \in \langle \mathcal{B} \rangle$, $E(X)$ is replaced by the adjusted expectation, $E_{\mathcal{D}}(X)$, and $Cov(X, Y)$ is replaced by the adjusted covariance, $Cov_{\mathcal{D}}(X, Y)$. $E_{\mathcal{D}}(X)$ may be viewed as the orthogonal projection of X into $[\mathcal{D}]$, and $Var_{\mathcal{D}}(X)$ the squared orthogonal distance from X to $[\mathcal{D}]$. If $\mathcal{B} = \{\mathcal{B}_1, \ldots, \mathcal{B}_r\}$ and $\mathcal{D} = \{\mathcal{D}_1, \ldots, \mathcal{D}_s\}$ then we may compute

$$E_{\mathcal{D}}(\mathcal{B}) = E(\mathcal{B}) + Cov(\mathcal{B}, \mathcal{D})Var^{\dagger}(\mathcal{D})(\mathcal{D} - E(\mathcal{D})); \quad (5)$$

$$Var_{\mathcal{D}}(\mathcal{B}) = Var(\mathcal{B}) - Cov(\mathcal{B}, \mathcal{D})Var^{\dagger}(\mathcal{D})Cov(\mathcal{D}, \mathcal{B}), \quad (6)$$

where $\mathcal{A}^{\dagger}$ represents the Moore-Penrose generalised inverse of A. For typical elements $X = b_0 + b^T\mathcal{B}$, $Y = \tilde{b}_0 + \tilde{b}^T\mathcal{B} \in \langle \mathcal{B} \rangle$, we have $E_{\mathcal{D}}(X) = b_0 + b^T E_{\mathcal{D}}(\mathcal{B})$ and $Cov_{\mathcal{D}}(X, Y) = b^T Var_{\mathcal{D}}(\mathcal{B})\tilde{b}$. We view this revision of beliefs as the replacement of $[\mathcal{B}]$ by the adjusted covariance, which we denote $[\mathcal{B}/\mathcal{D}_j]$.

Suppose that we are willing to give a full probability specification for our data, $\mathcal{D}$, and also the full joint distribution between $\mathcal{B}^* \subset \mathcal{B}$ and $\mathcal{D}$. Then, we are able to to construct $[\mathcal{B}^*_P|\mathcal{D}]$ but not $[\mathcal{B}_P|\mathcal{D}]$. Even if we can construct $[\mathcal{B}/\mathcal{D}_P]$, this update does not utilise the additional information we obtain about $\mathcal{B}^*$ from the full specification between $\mathcal{B}^*$ and $\mathcal{D}$. In the next section we clarify the types of models we are interested in which have differing levels of prior specifications and then go on to suggest methods to incorporate all the prior information into our belief revisions.

4 Augmented Bayes linear/Bayes graphical models

Our approach centres around exploiting structure within $\mathcal{B}$ to help us incorporate different amounts of prior information into our belief revision. In particular, we are thinking about structure that can be examined by a generalised conditional independence (g.c.i.) property,

$(\cdot \perp\!\!\!\perp \cdot)|\cdot$. The g.c.i. property was introduced by Smith (1989) as an extension of the work on probabilistic independence by Dawid (1979, 1980).

A collection of g.c.i. statements may be represented as a graphical model. A graphical model is a directed graph $\mathcal{G} = (V, E)$. $V = \{\mathcal{A}_1, \ldots, \mathcal{A}_t\}$ is the collection of nodes, where each $\mathcal{A}_i$ represents a collection of random quantities. E represents the collection of edges. If $(\mathcal{A}_i, \mathcal{A}_j) \in E$ then there is a directed arc from $\mathcal{A}_i$ to $\mathcal{A}_j$. We term $\mathcal{A}_i$ a parent of $\mathcal{A}_j$ and $\mathcal{A}_j$ a child of $\mathcal{A}_i$. The collection $\{\mathcal{A}_i : (\mathcal{A}_i, \mathcal{A}_j) \in E\}$ is the collection of parents of $\mathcal{A}_j$, written $pa(\mathcal{A}_j)$. The collection $\{\mathcal{A}_j : (\mathcal{A}_i, \mathcal{A}_j) \in E\}$ is the collection of children of $\mathcal{A}_i$, written $ch(\mathcal{A}_i)$.

Definition 1 *The directed acyclic graph, $\mathcal{G} = (V, E)$, is a g.c.i. graphical model if, for any $\mathcal{A}_i$, $\mathcal{A}_j \in V$, we have $(\mathcal{A}_i \perp\!\!\!\perp \mathcal{A}_j)|pa(\mathcal{A}_i)$, unless $\mathcal{A}_j$ is a descendent of $\mathcal{A}_i$.*

The most familiar example of a g.c.i. graphical model is the Bayesian graphical model (see for example, Pearl (1988), Lauritzen (1996), Cowell *et al.* (1999)). Here, the g.c.i. relation is taken to be probabilistic conditional independence, denoted $(\cdot \perp\!\!\!\perp_P \cdot)|\cdot$. For a general node, $\mathcal{A}_i$, we specify the conditional density function $f(a_i|pa(a_i))$ if $pa(\mathcal{A}_i) \neq \emptyset$. If $pa(\mathcal{A}_i) = \emptyset$ then the marginal density function, $f(a_i)$, of $\mathcal{A}_i$ is specified.

Goldstein & Wilkinson (2000) introduced a further example of a g.c.i. graphical model: the Bayes linear graphical model. The g.c.i. relation is taken to be belief separation. If $[\mathcal{X}]$, $[\mathcal{Y}]$, $[\mathcal{Z}]$ are three belief structures, then we say that $\mathcal{Z}$ separates $\mathcal{X}$ from $\mathcal{Y}$, written $(\mathcal{X} \perp\!\!\!\perp_S \mathcal{Y})|\mathcal{Z}$, if for all $X \in [\mathcal{X}]$, $Y \in [\mathcal{Y}]$,

$$Cov_{\mathcal{Z}}(X, Y) \quad = \quad 0. \tag{7}$$

Goldstein (1990) showed that belief separation was a g.c.i. statement. If $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$ are finite then equation (7) may be expressed as the requirement that

$$Cov(\mathcal{X}, \mathcal{Y}) \quad = \quad Cov(\mathcal{X}, \mathcal{Z})Var^{\dagger}(\mathcal{Z})Cov(\mathcal{Z}, \mathcal{Y}). \tag{8}$$

On a Bayes linear graphical model, for a general node, $\mathcal{A}_i$, we specify $E(\mathcal{A}_i)$, $Var(\mathcal{A}_i)$ and $Cov(\mathcal{A}_i, pa(\mathcal{A}_i))$. Goldstein & Wilkinson (2000; p313) show that it is equivalent to specify $E_{pa(\mathcal{A}_i)}(\mathcal{A}_i)$ and $Var_{pa(\mathcal{A}_i)}(\mathcal{A}_i)$ drawing a direct comparison between the specifications made in a Bayes and Bayes linear graphical model.

In this paper, we introduce a third type, which we term an augmented Bayes linear/Bayes graphical model.

Definition 2 *An augmented Bayes linear/Bayes (BLB) graphical model, $\mathcal{G} = (V, E)$, is a directed acyclic graph whose nodes, V, may be separated into two collections, $\mathcal{B} = \{B_1, \ldots, B_r\}$ and $\mathcal{D} = \{\mathcal{D}_1, \ldots, \mathcal{D}_s\}$. Certain nodes are joined by directed arcs, restricted to the three conditions*

1. *For any $\mathcal{D}_j \in \mathcal{D}$ there is a single node $\mathcal{B}_j \in \mathcal{B}$ such that $pa(\mathcal{D}_j) = \mathcal{B}_j$.*

2. *For any $\mathcal{D}_j \neq \mathcal{D}_k$ we have $pa(\mathcal{D}_j) \cap pa(\mathcal{D}_k) = \emptyset$.*

3. *For any $\mathcal{D}_j \in \mathcal{D}$, we have $ch(\mathcal{D}_j) = \emptyset$.*

For each $\mathcal{B}_i \in \mathcal{B}$ we specify $E(\mathcal{B}_i)$, $Var(\mathcal{B}_i)$ and $Cov(\mathcal{B}_i, pa(\mathcal{B}_i))$. For each $\mathcal{D}_j \in \mathcal{D}$ we specify $f(d_j, pa(d_j))$ and let $\mathcal{D}_{j,P}$ be the collection of nodes required to express $f(d_j)$ as a belief structure and set $\mathcal{D}_P = \cup_{j=1}^{s} \mathcal{D}_{j,P}$. For any nodes $\mathcal{A}_i$, $\mathcal{A}_j \in \mathcal{B} \cup \mathcal{D}_P$ we assert $(\mathcal{A}_i \perp\!\!\!\perp_S \mathcal{A}_j)|pa(\mathcal{A}_i)$, unless $\mathcal{A}_j$ is a descendent of $\mathcal{A}_i$.

Consider a node $\mathcal{D}_j \in \mathcal{D}$ with $pa(\mathcal{D}_j) = \mathcal{B}_j$. In the augmented BLB graphical model, the full joint probability distribution of $\mathcal{D}_j$ and $\mathcal{B}_j$ is specified and the implied belief structure has base $\mathcal{B}_{j,P} \cup \mathcal{D}_{j,P}$. The separation statements of Definition 2 allow us to evaluate $Cov(X, Y)$ for all $X \in \langle \mathcal{B} \rangle$, $Z \in \langle \mathcal{D}_{j,P} \rangle$. However, we cannot evaluate $Cov(X, Z)$ for all $X \in \langle \mathcal{B} \rangle$, $Z \in \langle \mathcal{B}_{j,P} \rangle$ since, for example, we only directly specify $Cov(\mathcal{B}_j, pa(\mathcal{B}_j))$ and so cannot construct the belief structure with base $\mathcal{B}_{j,P} \cup pa(\mathcal{B}_j)$. The belief structure with base $\mathcal{B} \cup \mathcal{D}_P$ is fully specified but there is no other random quantity, which is linearly independent with this collection and whose mean and variance we have specified in Definition 2, which we can add to this base and maintain a fully specified belief-structure. In particular, notice that $[\mathcal{B}_{j,P} \cup \mathcal{D}_{j,P}]$ is not a subspace of $[\mathcal{B} \cup \mathcal{D}_P]$.

5 Revising our beliefs following the observation of a single node

We are interested in revising our beliefs over $\mathcal{B}$ following observation of nodes in the collection $\mathcal{D}$. Initially, we restrict attention to the observation of a single node $\mathcal{D}_j \in \mathcal{D}$ and our proposed method of revising our expectations and variances over $\mathcal{B}$ is given in Definition 3.

Definition 3 *In an augmented BLB graphical model we define the revised expectation of any $X \in \langle \mathcal{B} \rangle$ following the observation of $\mathcal{D}_j$ to be the quantity*

$$E(X||\mathcal{D}_j) \quad = \quad E(E_{\mathcal{B}_j}(X)|\mathcal{D}_j), \tag{9}$$

where $\mathcal{B}_j = pa(\mathcal{D}_j)$. For any X, $Y \in \langle \mathcal{B} \rangle$ we define the revised covariance of X and Y following the observation of $\mathcal{D}_j$ to be the quantity

$$Cov(X, Y||\mathcal{D}_j) \quad = \quad Cov_{\mathcal{B}_j}(X, Y) + Cov(E_{\mathcal{B}_j}(X), E_{\mathcal{B}_j}(Y)|\mathcal{D}_j). \tag{10}$$

Our motivation for the revisions given in Definition 3 is intuitive rather than formal. Suppose we consider adjusting our beliefs using Bayes linear methods. Following the observation of $\mathcal{D}_{j,P}$ for any $X \in \langle \mathcal{B} \rangle$ we assign the adjusted expectation $E_{\mathcal{D}_{j,P}}(X)$ and for any $X, Y \in \langle \mathcal{B} \rangle$ we assign the adjusted covariance $Cov_{\mathcal{D}_{j,P}}(X, Y)$. Notice that since $ch(\mathcal{D}_j) = \emptyset$ then we have $(\mathcal{B} \perp\!\!\!\perp_S \mathcal{D}_{j,P})|\mathcal{B}_j$ and so we can exploit local computation to calculate our adjusted quantities. Utilising equations (16) and (17) of Goldstein & Wilkinson (2000), we have, for any $X, Y \in \langle \mathcal{B} \rangle$,

$$\begin{aligned} E_{\mathcal{D}_{j,P}}(X) \quad &= \quad E_{\mathcal{D}_{j,P}}(E_{\mathcal{B}_j}(X)); &(11)\\ Cov_{\mathcal{D}_{j,P}}(X, Y) \quad &= \quad Cov_{\mathcal{B}_j}(X, Y) + Cov_{\mathcal{D}_{j,P}}(E_{\mathcal{B}_j}(X), E_{\mathcal{B}_j}(Y)). &(12) \end{aligned}$$

We may calculate our adjusted quantities by performing calculations between $\mathcal{D}_{j,P}$ and $\mathcal{B}_j$ and then between $\mathcal{B}_j$ and $\mathcal{B}$. However, the calculation involving $\mathcal{D}_{j,P}$ and $\mathcal{B}_j$ does not fully utilise the full probability specification between $\mathcal{D}_j$ and $\mathcal{B}_j$. In particular, we have additional information about the reliability of $E_{\mathcal{D}_j,P}(\mathcal{B}_j)$ given by $Var(\mathcal{B}_j|\mathcal{D}_j)$. In a similar spirit to the variance modified linear Bayes estimator developed by Goldstein (1979, 1983), a natural generalisation of equations (11) and (12) to incorporate the benefits of performing a full posterior analysis of $\mathcal{B}_j$ given $\mathcal{D}_j$ is to use Bayes conditioning for the calculations between $\mathcal{D}_{j,P}$ and $\mathcal{B}_j$ rather than Bayes linear calculations. This leads directly to equations (9) and (10). Notice how if $X \in \langle \mathcal{B}_j \rangle$ then $E(X \parallel \mathcal{D}_j) = E(X|\mathcal{D}_j)$ and if $X, Y \in \langle \mathcal{B}_j \rangle$ then $Cov(X, Y \parallel \mathcal{D}_j) = Cov(X, Y|\mathcal{D}_j)$.

An alternative interpretation is to notice that since for all $X \in \langle \mathcal{B} \rangle$, $E_{\mathcal{D}_{j,P}}(X)$ is the orthogonal projection of X into $[\mathcal{D}_{j,P}]$ and $[\mathcal{D}_{j,P}]$ is the belief structure representation of a full probability measure then, see for example Cambanis (1982), $E_{\mathcal{D}_{j,P}}(X)$ is also the conditional expectation of X given $\mathcal{D}_j$, that is

$$E_{\mathcal{D}_{j,P}}(X) \;=\; E(X|\mathcal{D}_j) \;=\; E(X \parallel \mathcal{D}_j). \tag{13}$$

Thus, $E(X \parallel \mathcal{D}_j)$ is the conditional expectation for all $X \in \langle\mathcal{B}\rangle$. It will not, in general, be the case that $E_{\mathcal{B}_j}(X) = E(X|\mathcal{B}_j)$. Now, equation (11) may be obtained by decomposing X into two orthogonal components,

$$X \;=\; E_{\mathcal{B}_j}(X) + \{X - E_{\mathcal{B}_j}(X)\}, \tag{14}$$

and observing that the assertion that $(\mathcal{B} \perp\!\!\!\perp_S \mathcal{D}_{j,P})|\mathcal{B}_j$ implies that for all $X \in \langle\mathcal{B}\rangle$, $Y \in \langle\mathcal{D}_{j,P}\rangle$, $Cov(X - E_{\mathcal{B}_j}(X), Y) = 0$. If a full probability specification had been made between $\mathcal{B}$ and $\mathcal{D}_j$ then we could use equation (14) to calculate $Var(X|\mathcal{D}_j)$ as

$$\begin{aligned} Var(X|\mathcal{D}_j) &= Var(E_{\mathcal{B}_j}(X)|\mathcal{D}_j) + Var(X - E_{\mathcal{B}_j}(X)|\mathcal{D}_j) + \\ &\qquad 2Cov(E_{\mathcal{B}_j}(X), X - E_{\mathcal{B}_j}(X)|\mathcal{D}_j) \qquad (15) \\ &= Var(E_{\mathcal{B}_j}(X)|\mathcal{D}_j) + E(\{X - E_{\mathcal{B}_j}(X)\}^2|\mathcal{D}_j) + \\ &\qquad 2E(E_{\mathcal{B}_j}(X)\{X - E_{\mathcal{B}_j}(X)\}|\mathcal{D}_j), \qquad (16) \end{aligned}$$

where equation (16) follows from equation (15) by applying equation (13). Now if $X \notin \langle\mathcal{B}_j\rangle$ then our specification is not sufficiently detailed to allow us to compute $E(\{X - E_{\mathcal{B}_j}(X)\}^2|\mathcal{D}_j)$ and $E(E_{\mathcal{B}_j}(X)\{X - E_{\mathcal{B}_j}(X)\}|\mathcal{D}_j)$. However, observe that

$$\begin{aligned} E(E(\{X - E_{\mathcal{B}_j}(X)\}^2|\mathcal{D}_j)) &= E(\{X - E_{\mathcal{B}_j}(X)\}^2) \qquad (17) \\ &= Var_{\mathcal{B}_j}(X), \qquad (18) \end{aligned}$$

and

$$\begin{aligned} E(E(E_{\mathcal{B}_j}(X)\{X - E_{\mathcal{B}_j}(X)\}|\mathcal{D}_j)) &= E(E_{\mathcal{B}_j}(X)\{X - E_{\mathcal{B}_j}(X)\}) \qquad (19) \\ &= Cov(E_{\mathcal{B}_j}(X), X - E_{\mathcal{B}_j}(X)) \qquad (20) \\ &= 0, \qquad (21) \end{aligned}$$

so we can calculate the expectations of $E(\{X - E_{\mathcal{B}_j}(X)\}^2|\mathcal{D}_j)$ and $E(E_{\mathcal{B}_j}(X)\{X - E_{\mathcal{B}_j}(X)\}|\mathcal{D}_j)$ from our prior specifications. A plausible approximation to $Var(X|\mathcal{D}_j)$ is $Var(E_{\mathcal{B}_j}(X)|\mathcal{D}_j) + Var_{\mathcal{B}_j}(X)$. We could view equation (10) as an approximation to $Cov(X, Y|\mathcal{D}_j)$ given only a limited specification between $\mathcal{B}$ and $\mathcal{D}_j$.

We may consider $Cov(X, Y \parallel \mathcal{D}_j)$ to be either a natural generalisation of $Cov_{\mathcal{D}_{j,P}}(X, Y)$ to incorporate a more detailed partial prior specification, given only a partial prior specification, an approximation to $Var(X, Y|\mathcal{D}_j)$. Observe that

$$\begin{aligned} E(Cov(E_{\mathcal{B}_j}(X), E_{\mathcal{B}_j}(Y)|\mathcal{D}_j)) &= \\ &\quad E(E_{\mathcal{B}_j}(X)E_{\mathcal{B}_j}(Y)) - E(E_{\mathcal{D}_{j,P}}(E_{\mathcal{B}_j}(X))E_{\mathcal{D}_{j,P}}(E_{\mathcal{B}_j}(Y))) \qquad (22) \\ &= E(E_{\mathcal{B}_j}(X) - E_{\mathcal{D}_{j,P}}(E_{\mathcal{B}_j}(X))E_{\mathcal{B}_j}(Y) - E_{\mathcal{D}_{j,P}}(E_{\mathcal{B}_j}(Y))) \qquad (23) \\ &= Cov_{\mathcal{D}_{j,P}}(E_{\mathcal{B}_j}(X), E_{\mathcal{B}_j}(Y)), \qquad (24) \end{aligned}$$

where equation (23) follows from the fact that since $E_{\mathcal{D}_{j,P}}$ is the projection operator, for any $V, W \in \langle\mathcal{B}\rangle$, $Cov(V - E_{\mathcal{D}_{j,P}}(V), E_{\mathcal{D}_{j,P}}(W)) = 0$. Taking the expectation of equation (10), and using equation (24) we find that

$$E(Cov(X, Y|\mathcal{D}_j)) \;=\; E(Cov(X, Y \parallel \mathcal{D}_j)) \;=\; Cov_{\mathcal{D}_{j,P}}(X, Y). \tag{25}$$

We view the belief revision in Definition 3 as for each $X, Y \in \langle \mathcal{B} \rangle$ replacing $E(X)$ by $E(X \parallel \mathcal{D}_j)$ and $Cov(X,Y)$ by $Cov(X, Y \parallel \mathcal{D}_j)$. Our prior beliefs were collected into the belief structure $[\mathcal{B}]$ and we collect our revised beliefs together as the belief structure $[\mathcal{B} \parallel \mathcal{D}_j]$.

The revision of our prior specifications over the linear space $\langle \mathcal{B} \rangle$ have been defined in Definition 3 through the use of the separation statement $(\mathcal{B} \perp\!\!\!\perp_S \mathcal{D}_{j,P})|\mathcal{B}_j$. We now investigate whether this definition is consistent in the sense that if collections $\mathcal{A} = \{\mathcal{A}_1, \ldots, \mathcal{A}_r\}$, $\mathcal{C} = \{\mathcal{C}_1, \ldots, \mathcal{C}_s\} \subset \mathcal{B}$ are such that $(\mathcal{A} \perp\!\!\!\perp_S \mathcal{D}_{j,P})|\mathcal{C}$ can we perform local computations so that the revision of $\mathcal{A}$ by $\mathcal{D}_{j,P}$ is obtained via calculations between $\mathcal{D}_{j,P}$ and $\mathcal{C}$ and between $\mathcal{C}$ and $\mathcal{A}$? If so, how are these calculations performed? Also, we wish to explore when belief separation is preserved following the revision of belief. We shall make use of the following lemma; the proof is in the appendix.

Lemma 1 *Let $\mathcal{A} = \{\mathcal{A}_1, \ldots, \mathcal{A}_r\}$, $\mathcal{C} = \{\mathcal{C}_1, \ldots, \mathcal{C}_s\}$, $\mathcal{F} = \{\mathcal{F}_1, \ldots, \mathcal{F}_t\}$ be three collections of random quantities from $\mathcal{B}$.*

1. *If $(\mathcal{A} \perp\!\!\!\perp_S \mathcal{D}_{j,P})|\mathcal{C}$ then $(\mathcal{A} \perp\!\!\!\perp_S \mathcal{B}_j)|\mathcal{C}$, where $\mathcal{B}_j = pa(\mathcal{D}_j)$.*
2. *If $(\mathcal{A} \perp\!\!\!\perp_S \mathcal{F})|\mathcal{C} \cup \mathcal{D}_{j,P}$ then either $(\mathcal{A} \perp\!\!\!\perp_S \mathcal{D}_{j,P})|\mathcal{C}$ or $(\mathcal{F} \perp\!\!\!\perp_S \mathcal{D}_{j,P})|\mathcal{C}$.*

The following theorem shows that we can perform local computations to obtain the revision of belief; the proof is in the appendix.

Theorem 1 *Suppose collections $\mathcal{A} = \{\mathcal{A}_1, \ldots, \mathcal{A}_r\}$, $\mathcal{C} = \{\mathcal{C}_1, \ldots, \mathcal{C}_s\} \subset \mathcal{B}$ are such that $(\mathcal{A} \perp\!\!\!\perp_S \mathcal{D}_{j,P})|\mathcal{C}$, then:*

$$\begin{aligned} (i) \quad E(\mathcal{A} \parallel \mathcal{D}_j) &= E(E_{\mathcal{C}}(\mathcal{A}) \parallel \mathcal{D}_j); & (26) \\ (ii) \quad Var(\mathcal{A} \parallel \mathcal{D}_j) &= Var_{\mathcal{C}}(\mathcal{A}) + Var(E_{\mathcal{C}}(\mathcal{A}) \parallel \mathcal{D}_j); & (27) \\ (iii) \quad Cov(\mathcal{C}, \mathcal{A} \parallel \mathcal{D}_j) &= Cov(\mathcal{C}, E_{\mathcal{C}}(\mathcal{A}) \parallel \mathcal{D}_j); & (28) \end{aligned}$$

Theorem 1 is the analogous theorem to Theorem 4 of Goldstein & Wilkinson (2000). Theorem 1 allows us to make use of the separation statement $(\mathcal{A} \perp\!\!\!\perp_S \mathcal{D}_{j,P})|\mathcal{C}$ to perform a revision of belief step between $\mathcal{D}_j$ and $\mathcal{C}$, a collection of beliefs for whom the path between them on the moral graph contains a mixture of fully specified and partially specified nodes since $(\mathcal{C} \perp\!\!\!\perp_S \mathcal{D}_{j,P})|\mathcal{B}_j$, and a Bayes linear adjustment between $\mathcal{C}$ and $\mathcal{A}$, a pair with only a partial specification made between them. We now consider when separations in $[\mathcal{B}]$ are preserved in $[\mathcal{B} \parallel \mathcal{D}_j]$. We have the following theorem; the proof is in the appendix.

Theorem 2 *Suppose we have three collections, $\mathcal{A} = \{\mathcal{A}_1, \ldots, \mathcal{A}_r\}$, $\mathcal{C} = \{\mathcal{C}_1, \ldots, \mathcal{C}_s\}$, $\mathcal{F} = \{\mathcal{F}_1, \ldots, \mathcal{F}_t\} \subset \mathcal{B}$ such that $(\mathcal{A} \perp\!\!\!\perp_S \mathcal{F})|\mathcal{C}$ and that we denote the respective revised (by $\mathcal{D}_j$) belief structures as $[\mathcal{A} \parallel \mathcal{D}_j]$, $[\mathcal{C} \parallel \mathcal{D}_j]$, $[\mathcal{F} \parallel \mathcal{D}_j]$. If, additionally, $(\mathcal{A} \perp\!\!\!\perp_S \mathcal{F})|\mathcal{C} \cup \mathcal{D}_{j,P}$ then $([\mathcal{A} \parallel \mathcal{D}_j] \perp\!\!\!\perp_S [\mathcal{F} \parallel \mathcal{D}_j])|[\mathcal{C} \parallel \mathcal{D}_j]$, that is*

$$Cov(\mathcal{A}, \mathcal{F} \parallel \mathcal{D}_j) = Cov(\mathcal{A}, \mathcal{C} \parallel \mathcal{D}_j) Var^{\dagger}(\mathcal{C} \parallel \mathcal{D}_j) Cov(\mathcal{C}, \mathcal{F} \parallel \mathcal{D}_j). \quad (29)$$

Notice from Theorem 1 of Goldstein & Wilkinson (2000) that if we view our augmented BLB graphical model as a Bayes linear graphical model, then $([\mathcal{A}/\mathcal{D}_{j,P}] \perp\!\!\!\perp_S [\mathcal{F}/\mathcal{D}_{j,P}])|[\mathcal{C}/\mathcal{D}_{j,P}]$. The belief structure $[\mathcal{B} \parallel \mathcal{D}_j]$ preserves the same separations as $[\mathcal{B}/\mathcal{D}_{j,P}]$. We consider our modification to be the replacement of $[\mathcal{B}/\mathcal{D}_{j,P}]$ by $[\mathcal{B} \parallel \mathcal{D}_j]$, a structure which incorporates the additional information provided by the full specification between $\mathcal{D}_j$ and $\mathcal{B}_j$ but maintains the same separations as the Bayes linear adjustment. If we additionally wanted to observe quantities in $\mathcal{B}$, we may perform Bayes linear adjustments on the structure $[\mathcal{B} \parallel \mathcal{D}_j]$. Suppose we wish to observe the collection $\mathcal{C} = \{\mathcal{C}_1, \dots, \mathcal{C}_s\} \subset \mathcal{B}$ and perform Bayes linear adjustment of $[\mathcal{B} \parallel \mathcal{D}_j]$ following the observation of $\mathcal{C}$, where we have also revised our beliefs about $\mathcal{C}$ by $\mathcal{D}_j$. We denote the adjusted mean and variance for any X in $\langle \mathcal{B} \rangle$ by $E_{\mathcal{C}\|\mathcal{D}_j}(X)$ and $Var_{\mathcal{C}\|\mathcal{D}_j}(X)$. We have the following corollary to Theorem 1.

Corollary 1 *Suppose collections* $\mathcal{A} = \{\mathcal{A}_1, \dots, \mathcal{A}_r\}$, $\mathcal{C} = \{\mathcal{C}_1, \dots, \mathcal{C}_s\} \subset \mathcal{B}$ *are such that* $(\mathcal{A} \perp\!\!\!\perp_S \mathcal{D}_{j,P})|\mathcal{C}$, *then:*

$$\begin{aligned}
\textit{(i)} \quad E_{\mathcal{C}\|\mathcal{D}_j}(\mathcal{A}) &= E(\mathcal{A} \parallel \mathcal{D}_j) + Cov(\mathcal{A}, \mathcal{C} \parallel \mathcal{D}_j) Var^{\dagger}(\mathcal{C} \parallel \mathcal{D}_j)(\mathcal{C} - E(\mathcal{C} \parallel \mathcal{D}_j) \\
&= E_{\mathcal{C}}(\mathcal{A}); \qquad (30) \\
\textit{(ii)} \quad Var_{\mathcal{C}\|\mathcal{D}_j}(\mathcal{A}) &= Var(\mathcal{A} \parallel \mathcal{D}_j) - Cov(\mathcal{A}, \mathcal{C} \parallel \mathcal{D}_j) Var^{\dagger}(\mathcal{C} \parallel \mathcal{D}_j) Cov(\mathcal{C}, \mathcal{A} \parallel \mathcal{D}_j) \\
&= Var_{\mathcal{C}}(\mathcal{A}). \qquad (31)
\end{aligned}$$

Corollary 1 shows that $[[\mathcal{A} \parallel \mathcal{D}_j]/[\mathcal{C} \parallel \mathcal{D}_j]] = [\mathcal{A}/\mathcal{C}]$ and is reassuring: if $(\mathcal{A} \perp\!\!\!\perp_S \mathcal{D}_{j,P})|\mathcal{C}$ then we would expect to gain no further information about $\mathcal{A}$ by revising by both $\mathcal{D}_j$ and $\mathcal{C}$ than if we just revised by $\mathcal{C}$; the relationship between $\mathcal{A}$ and $\mathcal{C}$ is Bayes linear. The modified revision of $\mathcal{B}$ by $\mathcal{D}_j$ rather the Bayes linear adjustment does not affect our understanding of the separations in the augmented BLB graphical model.

5.1 The introductory example revisited

The abstract motivational example of Section 1 may be viewed as an augmented BLB model with three nodes, $\mathcal{B}_1$, $\mathcal{B}_2$ and $\mathcal{D}_1$, and two directed arcs, $(\mathcal{B}_1, \mathcal{D}_1)$ and $(\mathcal{B}_1, \mathcal{B}_2)$. We may easily verify that $(\mathcal{B}_2 \perp\!\!\!\perp_S \mathcal{D}_{1,P})|\mathcal{B}_1$. Following the observation of $\mathcal{D}_1$, we revise our beliefs over $\mathcal{B} = \{\mathcal{B}_1, \mathcal{B}_2\}$ using equations (9) and (10). We find that

$$E(\mathcal{B}||\mathcal{D}_1) = \begin{pmatrix} \mathcal{D}_1 \\ \frac{1}{2} + \frac{1}{18}\mathcal{D}_1 \end{pmatrix}; \qquad (32)$$

$$Var(\mathcal{B}||\mathcal{D}_1) = \begin{pmatrix} 1 - \mathcal{D}_1^2 & \frac{1}{18} - \frac{1}{18}\mathcal{D}_1^2 \\ \frac{1}{18} - \frac{1}{18}\mathcal{D}_1^2 & \frac{1}{20} - \frac{1}{324}\mathcal{D}_1^2 \end{pmatrix}. \qquad (33)$$

Observe how our intuition outlined in Section 1 is captured. If $\mathcal{D}_1 = 0$ then $Var(\mathcal{B} \parallel \mathcal{D}_1 = 0) = Var(\mathcal{B})$, reflecting the uninformative nature of this observation. If $\mathcal{D}_1 = \pm 1$ then $Var(\mathcal{B}_1||\mathcal{D}_1 = \pm 1) = 0$ and $Cov(\mathcal{B}_1, \mathcal{B}_2||\mathcal{D}_1 = \pm 1) = 0$, whilst $Var(\mathcal{B}_2||\mathcal{D}_1 = \pm 1) = Var_{\mathcal{B}_1}(\mathcal{B}_1)$. In Section 1, we argued that an observation of $\mathcal{D}_1 = \pm 1$ was equivalent to observing both $\mathcal{D}_1$ and $\mathcal{B}_1$. The separability statement $(\mathcal{B}_2 \perp\!\!\!\perp_S \mathcal{D}_{1,P})|\mathcal{B}_1$ implies that $\mathcal{B}_1$ is sufficient for this revision and so once more our intuition is captured. The belief revision of the variances given by equation (10) allows us a systematic way of capturing our intuition without requiring further specification.

6 Partition testing

Software testing is performed to reduce the risk to the producer and the user of releasing a poor quality product. Typical quantities of interest concern the reliability of the software. Test managers have to determine what subset of all possible inputs should be selected for testing in order to maximise the quality of the software constrained by the cost and the time of testing. Rees *et al.* (2001) gives a discussion of software testing in industry, focusing upon the problems faced by test managers and the difficulties of testing. In partition testing, the input domain I is partitioned into k disjoint subdomains, $I_1, \ldots, I_k$ such that each member of the subdomain is judged to have the same effect upon the system. Typical [**REFS**] assumptions include the judgement that test results for inputs in one subdomain provide no information about the test results for inputs in any of the other subdomains and that, within the same subdomain, all inputs yield either all correct outputs or all incorrect outputs. Coolen *et al.* (2001) proposed the use of Bayes linear methods in partition testing. This approach does not require the two assumptions outlined above (although they can be modelled) and makes explicit use of the expert knowledge held by the software testers. Letting X_{gi} be the value of the ith input in the I_gth subdomain and $X_g = \{X_{g1}, X_{g2}, \ldots\}$ be the total collection of values for the potential inputs for the I_gth subdomain, then partition testing may be considered as the scenario where the collections $X_1, \ldots, X_k$ are judged to be co-exchangeable. As defined by Goldstein (1986a), this requires that the joint second-order specification for each X_g, X_h to be invariant under permutation; that is

$$E(X_{gi}) = \mu_g \ \forall g, i; \tag{34}$$

$$Var(X_{gi}) = d_g \ \forall g, i; \tag{35}$$

$$Cov(X_{gi}, X_{gj}) = c_{gg} \ \forall g, i \neq j; \tag{36}$$

$$Cov(X_{gi}, X_{hj}) = c_{gh} \ \forall g \neq h, i, j. \tag{37}$$

Assuming a potentially infinite number of available inputs in each subdomain, we may apply the representation theorem of Goldstein (1986a) to write

$$X_{gi} = \mathcal{M}(X_g) + \mathcal{R}_i(X_g), \tag{38}$$

where the $\mathcal{R}_i(X_g)$ are mutually uncorrelated with mean zero. They are also uncorrelated with each $\mathcal{M}(X_g)$. $\mathcal{M}(X_g)$ is constructed as

$$\mathcal{M}(X_g) = \lim_{n_g \to \infty} \frac{1}{n_g} \sum_{i=1}^{n_g} X_{gi}, \tag{39}$$

with the limit in mean square and so can be regarded as the underlying population mean for the I_gth subdomain. $\mathcal{R}_i(X_g)$ is then the discrepancy from $\mathcal{M}(X_g)$ for the ith input in I_g. Coolen *et al.* (2001) consider the situation of binary test results setting $X_{gi} = 0$ if the ith input in the I_gth subdomain gives the correct corresponding output and $X_{gi} = 1$ if the corresponding output is incorrect. In this simplification we have that $d_g = \mu_g(1 - \mu_g)$.

In this $0 - 1$ framework, $\mathcal{M}(X_g)$ may be viewed as the underlying average proportion of failures in the I_gth subdomain and so the collection of values $\mathcal{M}(X) = \{\mathcal{M}(X_1), \ldots, \mathcal{M}(X_k)\}$ is a fundamental collection of random quantities of interest, as linear combinations of the $\mathcal{M}(X_g)$ may be viewed as measures of the reliability of the software, for example, the overall total proportion of failures. Observe from the representation theorem, equation (38), and the fact that the $\mathcal{R}_i(X_g)$ are mutually uncorrelated, that if $\mathcal{M}(X_g)$ was known, each X_{gi} may be viewed as an independent Bernoulli trial with probability $\mathcal{M}(X_g)$ of being a failure; see Goldstein (1994; Section 14.2) for a discussion on the analogous situation of coin tossing. Thus, given $\mathcal{M}(X_g)$, we may think of the tests in the I_gth subdomain to be binomial sampling.

In the assertion of the co-exchangeable model, it is the covariances between each $\mathcal{M}(X_g)$ and $\mathcal{M}(X_h)$ that provides the ability to learn about inputs in other subdomains from observed inputs in a subdomain of interest. Typically, testers have a thorough knowledge about the effects of particular test cases but a weaker knowledge about the relationships between different test cases. This knowledge may be reflected by, having specified a co-exchangeable model, a willingness of the testers to expand their beliefs about the individual subdomains by expressing a full probability distribution for each $\mathcal{M}(X_g)$. If this is the case, we may think of the co-exchangeable model corresponding to an augmented BLB graphical model as we now explain.

Suppose that in the I_gth subdomain we consider performing n_g tests. The collection of these measurements $\mathcal{D}_{g,P} = \{X_{g1}, \ldots, X_{gn_g}\}$ form a partition and the co-exchangeable specification yields the separation statements, for each $g, h = 1, \ldots, k$, $g \neq h$, $(\mathcal{M}(X) \perp\!\!\!\perp_S \mathcal{D}_{g,P}) | \mathcal{M}(X_g)$ and $(\mathcal{D}_{h,P} \perp\!\!\!\perp_S \mathcal{D}_{g,P}) | \mathcal{M}(X_g)$. If, in addition to introducing the co-exchangeable model, we are willing to give a full probability specification for each $\mathcal{M}(X_g)$ then since tests in the I_gth subdomain, given $\mathcal{M}(X_g)$, may be viewed as binomial sampling, we have thus specified the full joint probability distribution between $\mathcal{M}(X_g)$ and $\mathcal{D}_{g,P}$ for each g. Thus, the collection of nodes $\mathcal{M}(X) \cup \{\cup_{g=1}^{k} \mathcal{D}_{g,P}\}$, with the separation statements, $(\mathcal{M}(X) \perp\!\!\!\perp_S \mathcal{D}_{g,P}) | \mathcal{M}(X_g)$ and $(\mathcal{D}_{h,P} \perp\!\!\!\perp_S \mathcal{D}_{g,P}) | \mathcal{M}(X_g)$ form an augmented BLB graphical model.

For simplicity of exposition, we shall assume that the prior distribution for each $\mathcal{M}(X_g)$ is Beta(α_g, β_g) as this is the standard conjugate prior distribution for binomial sampling. Performing n_g tests in the I_gth subdomain and observing x_g incorrect outputs gives the posterior distribution of $\mathcal{M}(X_g)$, $\mathcal{M}(X_g)|\mathcal{D}_g$, to be Beta$(\alpha_g + x_g, \beta_g + n_g - x_g)$. For software testing, we assume that, if a fault is found testing is halted and the fault repaired and a fresh test suite designed. Thus, we assume that all the observed test outcomes are correct, so that each $x_g = 0$.

In parallel with Coolen *et al.* (2001), we choose the triangle classification problem as an illustrative example. The input is three positive integers $A \geq B \geq C$ and the output should be whether the integers form the sides of a triangle and if they do, what kind of triangle they form. In Example 1, Coolen *et al.* (2001) consider a partition into five subdomains:

1. $(A < B) \vee (B < C)$ (illegal order)
2. $\begin{cases} (A \geq B + C) \wedge (A > B > C) & \text{(not a triangle)} \\ (A \geq B + C) \wedge (B = C) & \text{(not a triangle)} \end{cases}$
3. $\begin{cases} (A = B = C) & \text{(equilateral)} \\ (A = B > C) & \text{(isosceles)} \\ (A > B = C) \wedge (A < B + C) & \text{(isosceles)} \end{cases}$
4. $(A > B > C) \wedge (A^2 = B^2 + C^2)$ (right scalene)
5. $\begin{cases} (A > B > C) \wedge (A^2 < B^2 + C^2) & \text{(acute scalene)} \\ (A > B > C) \wedge (A^2 > B^2 + C^2) \wedge (A < B + C) & \text{(obtuse scalene)} \end{cases}$

These five subdomains are themselves a further partition of the nine subdomains taken by Weyuker & Ostrand (1980). We first specify our expectations and variances over the vector $\mathcal{M}(X) = (\mathcal{M}(X_1), \ldots, \mathcal{M}(X_5))^T$. We take the same prior specification as Coolen *et al.* (2001) and so assert $E(\mathcal{M}(X)) = (1/5,\ 1/5,\ 3/10,\ 2/5,\ 1/2)^T$ and

$$Var(\mathcal{M}(X)) = \begin{pmatrix} \frac{1}{125} & \frac{1}{500} & \frac{3}{1000} & \frac{1}{250} & \frac{1}{200} \\ \frac{1}{500} & \frac{1}{125} & \frac{3}{1000} & \frac{1}{250} & \frac{1}{200} \\ \frac{3}{1000} & \frac{1}{1000} & \frac{3}{200} & \frac{1}{250} & \frac{1}{200} \\ \frac{1}{250} & \frac{1}{250} & \frac{1}{250} & \frac{1}{50} & \frac{1}{100} \\ \frac{1}{200} & \frac{1}{200} & \frac{1}{200} & \frac{1}{100} & \frac{1}{40} \end{pmatrix} \tag{40}$$

From these we may fit the Beta distributions with the same means and variances. For example, $\mathcal{M}(X_1)$ is assigned the Beta(19/5, 76/5). Suppose that we consider observing tests in only a single subdomain. Using equations (9) and (10) we calculate our revised expectations and covariances. The revised expectations are equivalent to the Bayes linear expectations and so we shall not dwell upon them here. The revised variances may be expressed as follows:

$$Var^{-1}(\mathcal{M}(X) \parallel \mathcal{D}_1) = \frac{25n_1(646+58n_1+n_1^2)}{19(76+5n_1)}E_{5,1} + Var^{-1}(\mathcal{M}(X)); \tag{41}$$

$$Var^{-1}(\mathcal{M}(X) \parallel \mathcal{D}_2) = \frac{25n_2(646+58n_2+n_2^2)}{19(76+5n_2)}E_{5,2} + Var^{-1}(\mathcal{M}(X)); \tag{42}$$

$$Var^{-1}(\mathcal{M}(X) \parallel \mathcal{D}_3) = \frac{100n_3(273+40n_3+n_3^2)}{39(91+10n_3)}E_{5,3} + Var^{-1}(\mathcal{M}(X)); \tag{43}$$

$$Var^{-1}(\mathcal{M}(X) \parallel \mathcal{D}_4) = \frac{25n_4(165+34n_4+n_4^2)}{22(33+5n_4)}E_{5,4} + Var^{-1}(\mathcal{M}(X)); \tag{44}$$

$$Var^{-1}(\mathcal{M}(X) \parallel \mathcal{D}_5) = \frac{4n_5(81+28n_1+n_5^2)}{9(9+2n_5)}E_{5,5} + Var^{-1}(\mathcal{M}(X)); \tag{45}$$

where E_{5j} denotes the 5×5 matrix whose only non-zero entry is the (j,j)th entry, which is 1. The form of equations (41) - (45) is a direct consequence of separation statements $(\mathcal{M}(X) \perp\!\!\!\perp_S \mathcal{D}_{g,P}) | \mathcal{M}(X_g)$ for each $g = 1, \ldots, 5$. As we argued, we may view $Var(\mathcal{M}(X) \parallel \mathcal{D}_g)$ to be a sensible modification of $Var_{\mathcal{D}_{g,P}}(\mathcal{M}(X))$ to allow for the additional information provided by conditioning between $\mathcal{D}_g$ and $\mathcal{M}(X_g)$ and so we should compare the two variance matrices. Considering the first subdomain, we find:

$$Var(\mathcal{M}(X) \parallel \mathcal{D}_1) - Var_{\mathcal{D}_{1,P}}(\mathcal{M}(X)) = \frac{-19n_1(n_1+14)}{125(19+n_1)(n_1^2+39n_1+380)} \begin{pmatrix} 1 & \frac{1}{4} & \frac{3}{8} & \frac{1}{2} & \frac{5}{8} \\ \frac{1}{4} & \frac{1}{16} & \frac{3}{32} & \frac{1}{8} & \frac{5}{32} \\ \frac{3}{8} & \frac{3}{32} & \frac{9}{64} & \frac{3}{16} & \frac{15}{64} \\ \frac{1}{2} & \frac{1}{8} & \frac{3}{16} & \frac{1}{4} & \frac{5}{15} \\ \frac{5}{8} & \frac{5}{32} & \frac{15}{64} & \frac{5}{16} & \frac{25}{64} \end{pmatrix}, \tag{46}$$

so that for all sample sizes, $Var(\mathcal{M}(X) \parallel \mathcal{D}_1) < Var_{\mathcal{D}_{1,P}}(\mathcal{M}(X))$. Indeed, we may readily find that for all g and all n_g, $Var(\mathcal{M}(X) \parallel \mathcal{D}_g) < Var_{\mathcal{D}_{g,P}}(\mathcal{M}(X))$. This is not surprising. The Bayes linear adjusted variance effectively averages across all sample outcomes and so, since seeing all tests to be successes in a Binomial sample with the initial priors has a small probability, attaches little weight to the variance attached with the data outcome. However, in our revised variances, we can take the conditional variance $Var(\mathcal{M}(X_g)|\mathcal{D}_g)$ and use this to alter the weightings for the other $Var(\mathcal{M}(X)|\mathcal{D}_g)$.

7 Linking the revisions; belief transforms

Goldstein (1991) describes a general approach to the comparison of an original inner product over a collection, $\mathcal{B}$, over quantities of interest and a derived quadratic form over these random quantities.

Definition 4 *(Bachman & Narici (2000; Section 20.4)) A symmetric psd sesquilinear functional on a linear space $\langle \mathcal{B} \rangle$ is any real-valued functional, g, satisfying, for all X, Y, Z in*

$\langle \mathcal{B} \rangle$ *and numbers a, b,*

$$g(X,Y) \;=\; g(Y,X); \tag{47}$$
$$g(X,X) \;\geq\; 0; \tag{48}$$
$$g(aX+bY,Z) \;=\; ag(X,Z)+bg(Y,Z). \tag{49}$$

A real inner product space is an example of a symmetric psd sesquilinear functional; notice that, unlike a real inner product spaces, such functionals allow non-zero elements to have zero norm. Thus, if g relates to a covariance, we may have quantities in $\langle \mathcal{B} \rangle$ with zero variance. If there is a constant k for which $|g(X,Y)| \leq k \parallel X \parallel \parallel Y \parallel$ for all X, Y in $\langle \mathcal{B} \rangle$ then g is said to be bounded and k is the norm of the functional. Notice that, for observation of data $\mathcal{D}=d$, conditioning, $Cov(X,Y|\mathcal{D}=d)$, and Bayes linear adjustment, $Cov_{\mathcal{D}}(X,Y)$, are two examples of bounded symmetric psd sesquilinear functionals.

Theorem 3 *(Bachman & Narici (2000; Section 21.1)) A necessary and sufficient condition for g to be a bounded symmetric psd sesquilinear functional over the Hilbert space* $\langle \mathcal{B} \rangle$ *is that g is of the form*

$$g(X,Y) \;=\; (X,SY), \tag{50}$$

where S is a bounded self-adjoint operator over $\langle \mathcal{B} \rangle$*, with norm equal to the norm of g.*

Goldstein (1991) uses this construction in making the following definition.

Definition 5 *The bounded self-adjoint operator S defined by equation (50) is termed the belief transform for* $(\cdot,\cdot)$ *associated with g. The complementary belief transform T is defined by* $T=I-S$*, where I is the identity operator over* $\langle \mathcal{B} \rangle$*.*

Typically, we shall take $(\cdot,\cdot)$ to be the covariance inner product. If $\mathcal{B}$ is finite, consisting of r quantities $\mathcal{B}_i$, we construct S as follows. $(\cdot,\cdot)$ is represented by the matrix $Var(\mathcal{B})$. We form the $r \times r$ matrix U whose (i,j)th element is $g(\mathcal{B}_i,\mathcal{B}_j)$. The representation of S is $S(\mathcal{B}) = Var^{\dagger}(\mathcal{B})U$. For a general $Y=\tilde{b}_0+\tilde{b}^T\mathcal{B}$, we have $SY=\mathcal{B}^T S(\mathcal{B})\tilde{b}$.

The most familiar belief transform is in the Bayes linear setting when $g(X,Y)=Cov_{\mathcal{D}}(X, Y)$. The belief transform is denoted by $S_{\mathcal{D}}$ and the matrix representation of $S_{\mathcal{D}}$ is

$$S_{\mathcal{D}}(\mathcal{B}) \;=\; Var^{\dagger}(\mathcal{B})\{Var(\mathcal{B})-Cov(\mathcal{B},\mathcal{D})Var^{\dagger}(\mathcal{D})Cov(\mathcal{B},\mathcal{D})\}. \tag{51}$$

The complementary belief transform to $S_{\mathcal{D}}$ is termed the resolution transform, written $T_{\mathcal{D}}$. Goldstein (1981) introduces and studies the properties of $T_{\mathcal{D}}$, revealing that, for each $X \in \langle \mathcal{B} \rangle$, $T_{\mathcal{D}}(X)$ is the point in $[\mathcal{B}]$ which is closest to $E_{\mathcal{D}}(X)$ so that $T_{\mathcal{D}}(X)=E_{\mathcal{B}}(E_{\mathcal{D}}(X))$. $T_{\mathcal{D}}$ itself may be considered as a belief transform via the functional

$$RCov_{\mathcal{D}}(X,Y) \;=\; Cov(X,Y)-Cov_{\mathcal{D}}(X,Y) \;=\; Cov(X,T_{\mathcal{D}}(Y)). \tag{52}$$

When we want to explicitly designated the space over which $S_{\mathcal{D}}$ and $T_{\mathcal{D}}$ are operating, we write the respective transforms as $S_{\mathcal{B}/\mathcal{D}}$, $T_{\mathcal{B}/\mathcal{D}}$.

The importance of $S_{\mathcal{D}}$ is that it essentially transforms the original belief structure $[\mathcal{B}]$ into the adjusted belief structure $[\mathcal{B}/\mathcal{D}]$. In our revision of $[\mathcal{B}]$ into $[\mathcal{B} \parallel \mathcal{D}_j]$ in Section 5, the quadratic form for the revision is $Cov(X,Y \parallel \mathcal{D}_j)$, see equation (10). $Cov(X,Y \parallel \mathcal{D}_j=d_j)$ is the sum of two bounded symmetric psd sesquilinear functionals on $\langle \mathcal{B} \rangle$ and so is one itself. Thus, from Theorem 3, there is a bounded self-adjoint operator $S_{\mathcal{B}\parallel\mathcal{D}_j}$ satisfying, for each X, Y in $\langle \mathcal{B} \rangle$,

$$Cov(X,Y \parallel \mathcal{D}_j=d_j) \;=\; Cov(X,S_{\mathcal{B}\parallel\mathcal{D}_j}(Y)). \tag{53}$$

We term $S_{\mathcal{B}\|\mathcal{D}_j}$ the revised belief transform, with matrix representation $S_{\mathcal{B}\|\mathcal{D}_j}(\mathcal{B}) = Var^{\dagger}(\mathcal{B})$ $Var(\mathcal{B} \parallel \mathcal{D}_j = d_j)$. The complimentary revised belief transform is $T_{\mathcal{B}\|\mathcal{D}_j} = I - S_{\mathcal{B}\|\mathcal{D}_j}$, termed the revised resolution transform. Notice that whilst

$$RCov(X, Y \parallel \mathcal{D}_j = d_j) \;=\; Cov(X, Y) - Cov(X, Y \parallel \mathcal{D}_j = d_j) \;=\; Cov(X, T_{\mathcal{B}\|\mathcal{D}_j}(Y)), \quad (54)$$

it is possible that $RVar(X \parallel \mathcal{D}_j = d_j) < 0$ and so $RCov(X, Y \parallel \mathcal{D}_j = d_j)$ is not a symmetric psd sesquilinear functional.

$S_{\mathcal{B}\|\mathcal{D}_j}$ transforms the original belief structure into $[\mathcal{B} \parallel \mathcal{D}_j]$. We consider local computation properties of the transforms; we have the following corollary to Theorem 1.

Corollary 2 *Suppose collections* $\mathcal{A} = \{\mathcal{A}_1, \ldots, \mathcal{A}_r\}$, $\mathcal{C} = \{\mathcal{C}_1, \ldots, \mathcal{C}_s\} \subset \mathcal{B}$ *are such that* $(\mathcal{A} \perp\!\!\!\perp_S \mathcal{D}_{j,P})|\mathcal{C}$, *then for all* Y *in* $\langle \mathcal{A} \rangle$*:*

$$\begin{aligned} (i) \quad S_{\mathcal{A}\|\mathcal{D}_j}(Y) &= S_{\mathcal{A}/\mathcal{C}}(Y) + E_{\mathcal{A}}(S_{\mathcal{C}\|\mathcal{D}_j}(E_{\mathcal{C}}(Y))); && (55) \\ (ii) \quad T_{\mathcal{A}\|\mathcal{D}_j}(Y) &= E_{\mathcal{A}}(T_{\mathcal{C}\|\mathcal{D}_j}(E_{\mathcal{C}}(Y))). && (56) \end{aligned}$$

Notice that Corollary 2 permits the calculation of $S_{\mathcal{B}\|\mathcal{D}_j}$, $T_{\mathcal{B}\|\mathcal{D}_j}$ by setting $\mathcal{A} = \mathcal{B}$ and $\mathcal{C} = \mathcal{B}_j$.

We may calculation equations (9) and (10) by computing the pair $\{E(\mathcal{B} \parallel \mathcal{D}_j), S_{\mathcal{B}\|\mathcal{D}_j}(\mathcal{B})\}$. To calculate this pair, since $(\mathcal{B} \perp\!\!\!\perp_S \mathcal{D}_{j,P})|\mathcal{B}_j$, it is sufficient to calculate the evidence message

$$\{E(\mathcal{B}_j \parallel \mathcal{D}_j) = E(\mathcal{B}_j|\mathcal{D}_j), S_{\mathcal{B}_j\|\mathcal{D}_j}(\mathcal{B}_j) = Var^{\dagger}(\mathcal{B}_j)Var(\mathcal{B}_j|\mathcal{D}_j)\}. \quad (57)$$

Consider the Bayes linear adjustment of $\mathcal{B}$ following the observation of $\mathcal{D}_{j,P}$. To calculate the pair $\{E_{\mathcal{D}_{j,P}}(\mathcal{B}), S_{\mathcal{B}/\mathcal{D}_{j,P}}(\mathcal{B})\}$, equations (16) and (19) of Goldstein & Wilkinson (2000) show that it is sufficient to calculate the evidence message $\{E_{\mathcal{D}_{j,P}}(\mathcal{B}_j), S_{\mathcal{B}_j/\mathcal{D}_{j,P}}(\mathcal{B}_j)\}$. By comparing equation (26) with equation (16) of Goldstein & Wilkinson (2000) and equation (56) with equation (19) of Goldstein & Wilkinson (2000): the only difference between the calculation of $\{E(\mathcal{B} \parallel \mathcal{D}_j), S_{\mathcal{B}\|\mathcal{D}_j}(\mathcal{B})\}$ and $\{E_{\mathcal{D}_{j,P}}(\mathcal{B}), S_{\mathcal{B}/\mathcal{D}_{j,P}}(\mathcal{B})\}$ is in the evidence message received. Given the evidence message, they are computed in an identical fashion. The consistency of equations (9) and (10) allows us to calculate locally and once more, given the evidence message, these computations are identical to those in the Bayes linear adjustment.

8 Revising our beliefs following the observation of a data collection

Suppose we now consider revising our beliefs over $\mathcal{B}$ following observation of nodes in the collection $\mathcal{D}^* = \{\mathcal{D}_{(1)}, \ldots, \mathcal{D}_{(s')}\}$ where $\mathcal{D}^* \subseteq \mathcal{D}$. We assume that the nodes in $\mathcal{B}$ are suitably labelled so that $pa(\mathcal{D}_j) = \mathcal{B}_j$ and we let $\mathcal{D}^* = \cup_{j=1}^{s'} \mathcal{D}_{(j),P}$. In Section 5 we discussed the revision of $\mathcal{B}$ following the observation of a single node $\mathcal{D}_j$, see Definition 3. The proposed revision involved exploiting local computation techniques noting that $(\mathcal{B} \perp\!\!\!\perp_S \mathcal{D}_{j,P})|\mathcal{B}_j$. Theorem 1 and Corollary 2 illustrated that the revision given in Definition 3 provided a way to embed full conditioning into the Bayes linear model over $\mathcal{B}$ by utilising the separating node $\mathcal{B}_j = pa(\mathcal{D}_j)$. For the $\mathcal{D}^*$ notice that $(\mathcal{B} \perp\!\!\!\perp_S \mathcal{D}^*_P)|\mathcal{B}^*$, where $\mathcal{B}^* = \cup_{j=1}^{s'} \mathcal{B}_{(j)}$. This suggests a similar approach for the revision of $\mathcal{B}$ following the observation of $\mathcal{D}^*$: we perform the revision using calculations between the pair $\{\mathcal{D}^*, \mathcal{B}^*\}$ and between the pair the pair $\{\mathcal{B}^*, \mathcal{B}\}$. The pair $\{\mathcal{B}^*, \mathcal{B}\}$ are partially specified so that we may advocate a Bayes linear calculation between this pair; the pair $\{\mathcal{D}^*, \mathcal{B}^*\}$ involves a mixed level of specification and we consider a suitable revision of this pair.

Definition 6 *In an augmented BLB graphical model we define the revised expectation of any $X \in \langle \mathcal{B}^* \rangle$ following the observation of $\mathcal{D}^*$ to be the quantity*

$$E(X \parallel \mathcal{D}^*) = \sum_{j=1}^{s'} E(S^{\dagger}_{\mathcal{B}^* \parallel \mathcal{D}_{(j)}} S_{\mathcal{B}^* \parallel \mathcal{D}^*}(X) \parallel \mathcal{D}_{(j)}), \tag{58}$$

where

$$S_{\mathcal{B}^* \parallel \mathcal{D}^*}(X) = (\sum_{j=1}^{s'} S^{\dagger}_{\mathcal{B}^* \parallel \mathcal{D}_{(j)}} - (s-1)I)^{\dagger}(X). \tag{59}$$

For any X, $Y \in \langle \mathcal{B}^* \rangle$ we define the revised covariance of X and Y following the observation of $\mathcal{D}^*$ to be the quantity

$$Cov(X, Y \parallel \mathcal{D}^*) = Cov(X, S_{\mathcal{B}^* \parallel \mathcal{D}^*}(Y)). \tag{60}$$

To motivate this definition, consider three belief structures $[\mathcal{X}]$, $[\mathcal{Y}]$, $[\mathcal{Z}]$ with the property that $(\mathcal{X} \perp\!\!\!\perp_S \mathcal{Y})|\mathcal{Z}$. It can be shown that equivalent representations of equations (20) and (21) of Goldstein & Wilkinson (2000) are, for all $Z \in \langle \mathcal{Z} \rangle$,

$$S_{\mathcal{Z}/\mathcal{X} \cup \mathcal{Y}}(Z) = (S^{\dagger}_{\mathcal{Z}/\mathcal{X}} + S^{\dagger}_{\mathcal{Z}/\mathcal{Y}} - I)^{\dagger}(Z); \tag{61}$$

$$E_{\mathcal{Z}/\mathcal{X} \cup \mathcal{Y}}(Z) = E_{\mathcal{X}}(S^{\dagger}_{\mathcal{Z}/\mathcal{X}} S_{\mathcal{Z}/\mathcal{X} \cup \mathcal{Y}}(Z)) + E_{\mathcal{Y}}(S^{\dagger}_{\mathcal{Z}/\mathcal{Y}} S_{\mathcal{Z}/\mathcal{X} \cup \mathcal{Y}}(Z)). \tag{62}$$

Now, on the augmented BLB graphical model, if $\mathcal{D}^{\diamond}$, $\mathcal{D}^{\circ} \subset \mathcal{D}$ with $\mathcal{D}^{\diamond} \cap \mathcal{D}^{\circ} = \emptyset$ then $(\mathcal{D}^{\diamond}_P \perp\!\!\!\perp_S \mathcal{D}^{\circ}_P)|\mathcal{B}^*$. If we are performing the Bayes linear adjustment of $[\mathcal{B}^*]$ following the observation of $\mathcal{D}^*_P$, then we may adjust our beliefs via repeated use of equations (61) and (62) using any subsets of $\mathcal{D}^*$ we desire. In particular, setting $\mathcal{D}_{[k]} = \cup_{j=1}^{k} \mathcal{D}_{(j)}$ then we have $(\mathcal{D}_{[k],P} \perp\!\!\!\perp_S \mathcal{D}_{(k+1)})|\mathcal{B}^*$ for each $k = 2, \ldots, s'-1$ and so we may first merge the revisions by $\mathcal{D}_{(1)}$ and $\mathcal{D}_{(2)}$, and then merge the revisions by $\mathcal{D}_{[2]}$ and $\mathcal{D}_{(3)}$, and so on. By decomposing each $S_{\mathcal{B}^*/\mathcal{D}_{[k],P}}(\cdot)$ and $E_{\mathcal{D}_{[k],P}}(\cdot)$, we can see that we have $S_{\mathcal{B}^*/\mathcal{D}^*_P}(X) = (\sum_{j=1}^{s'} S^{\dagger}_{\mathcal{B}^*/\mathcal{D}_{(j),P}} - (s'-1)I)^{\dagger}(X)$ and $E_{\mathcal{D}^*_P}(X) = \sum_{j=1}^{s'} E_{\mathcal{D}^*_P}(S^{\dagger}_{\mathcal{B}^*/\mathcal{D}_{(j),P}} S_{\mathcal{B}^*/\mathcal{D}^*_P}(X))$ for any X in $\langle \mathcal{B}^* \rangle$. To perform this adjustment we only require the sets $\{E_{\mathcal{D}_{(j),P}}(\mathcal{B}^*), S_{\mathcal{B}^*/\mathcal{D}_{(j),P}}(\mathcal{B}^*)\}$ for each $j = 1, \ldots, s$. Noting that $(\mathcal{B}^* \perp\!\!\!\perp_S \mathcal{D}_{(j),P})|\mathcal{B}_{(j)}$, all that is required to calculate each $\{E_{\mathcal{D}_{(j),P}}(\mathcal{B}^*), S_{\mathcal{B}^*/\mathcal{D}_{(j),P}}(\mathcal{B}^*)\}$ is $[\mathcal{B}]$ and $\{E_{\mathcal{D}_{(j),P}}(\mathcal{B}_{(j)}), S_{\mathcal{B}_{(j)}/\mathcal{D}_{(j),P}}(\mathcal{B}_{(j)})\}$.

In Section 5 we showed that $[\mathcal{B} \parallel \mathcal{D}_{(j)}]$ has the same structure, in terms of its separations, as $[\mathcal{B}/\mathcal{D}_{(j),P}]$ but incorporates the additional information provided by the full probability specification between $\mathcal{D}_{(j)}$ and $\mathcal{B}_{(j)}$. Consequently, the same applies for the subspaces so that $[\mathcal{B}^* \parallel \mathcal{D}_{(j)}]$ has the same structure as $[\mathcal{B}^*/\mathcal{D}_{(j),P}]$. A natural way to incorporate this additional information at each pair $\{\mathcal{B}_{(j)}, \mathcal{D}_{(j)}\}$ in the revision of $[\mathcal{B}^*]$ by $\mathcal{D}^*$ is to replace each set $\{E_{\mathcal{D}_{(j),P}}(\mathcal{B}^*), S_{\mathcal{B}^*/\mathcal{D}_{(j),P}}(\mathcal{B}^*)\}$ by $\{E(\mathcal{B}^* \parallel \mathcal{D}_{(j)}), S_{\mathcal{B}^* \parallel \mathcal{D}}(\mathcal{B}^*)\}$. This leads to equations (58) - (60). Recall, from equation (13), that each $E(\mathcal{B}^* \parallel \mathcal{D}_{(j)}) = E_{\mathcal{D}_{(j),P}}(\mathcal{B}^*)$ so that, in a spirit similar to Goldstein (1979, 1983), equations (58) - (60) may be viewed as variance modified Bayes linear estimates for $\langle \mathcal{B}^* \rangle$.

Definition 7 *In an augmented BLB graphical model we define the revised expectation of any $X \in \langle \mathcal{B} \rangle$ following the observation of $\mathcal{D}^*$ to be the quantity*

$$E(X \| \mathcal{D}^*) = E(E_{\mathcal{B}^*}(X) \parallel \mathcal{D}^*), \tag{63}$$

For any X, $Y \in \langle \mathcal{B} \rangle$ we define the revised covariance of X and Y following the observation of $\mathcal{D}^*$ to be the quantity

$$Cov(X, Y \| \mathcal{D}^*) = Cov_{\mathcal{B}^*}(X, Y) + Cov(E_{\mathcal{B}^*}(X), E_{\mathcal{B}^*}(Y) \parallel \mathcal{D}^*). \tag{64}$$

Notice the similarity between Definition 3 and Definition 7. In Definition 3, we have $(\mathcal{B} \perp\!\!\!\perp_S \mathcal{D}_j)|\mathcal{B}_j$. For the Bayes linear adjustment of $[\mathcal{B}]$ by $\mathcal{D}_{j,P}$ it is sufficient to known $[\mathcal{B}]$ and the evidence pair $\{E_{\mathcal{D}_{j,P}}(\mathcal{B}_j),\ S_{\mathcal{B}_j/\mathcal{D}_{j,P}}(\mathcal{B}_j)\}$ and the revised adjustment of Definition 3 may be viewed as the modified Bayes linear adjustment, obtained by replacing the evidence pair to $\{E(\mathcal{B}_j \parallel \mathcal{D}_j),\ S_{\mathcal{B}_j\parallel\mathcal{D}_j}(\mathcal{B}_j)\}$: this evidence is absorbed into $\mathcal{B}$ using Bayes linear calculations. The same applies to the revision as described in Definition 6. We have $(\mathcal{B} \perp\!\!\!\perp_S \mathcal{D}^*_P)|\mathcal{B}^*$. The Bayes linear adjustment of $[\mathcal{B}]$ following observation of $\mathcal{D}^*_P$ requires only the evidence message $\{E_{\mathcal{D}^*_P}(\mathcal{B}^*),\ S_{\mathcal{B}^*/\mathcal{D}^*_P}(\mathcal{B}^*)\}$ and $[\mathcal{B}]$. We modify the evidence message to $\{E(\mathcal{B}^* \parallel \mathcal{D}^*),\ S_{\mathcal{B}^*\parallel\mathcal{D}^*}(\mathcal{B}^*)\}$ and absorb it into $[\mathcal{B}]$ using Bayes linear calculations. In a similar vein to Theorem 1 and Corollary 2 we may obtain the following theorem; the proof is similar to Theorem 1 noting that if $(\mathcal{A} \perp\!\!\!\perp_S \mathcal{D}^\circ_P)|\mathcal{C}$ then $(\mathcal{A} \perp\!\!\!\perp_S \mathcal{B}^\circ)|\mathcal{C}$.

Theorem 4 *Suppose collections* $\mathcal{A} = \{\mathcal{A}_1, \ldots, \mathcal{A}_r\}$, $\mathcal{C} = \{\mathcal{C}_1, \ldots, \mathcal{C}_s\} \subset \mathcal{B}$ *are such that* $(\mathcal{A} \perp\!\!\!\perp_S \mathcal{D}^*_P)|\mathcal{C}$, *then:*

$$\begin{aligned}
&\textit{(i)} & E(\mathcal{A} \parallel \mathcal{D}^*) &= E(E_{\mathcal{C}}(\mathcal{A}) \parallel \mathcal{D}^*); & (65)\\
&\textit{(ii)} & Var(\mathcal{A} \parallel \mathcal{D}^*) &= Var_{\mathcal{C}}(\mathcal{A}) + Var(E_{\mathcal{C}}(\mathcal{A}) \parallel \mathcal{D}^*); & (66)\\
&\textit{(iii)} & Cov(\mathcal{C}, \mathcal{A} \parallel \mathcal{D}^*) &= Cov(\mathcal{C}, E_{\mathcal{C}}(\mathcal{A}) \parallel \mathcal{D}^*); & (67)\\
&\textit{(iv)} & S_{\mathcal{A}\parallel\mathcal{D}^*}(Y) &= S_{\mathcal{A}/\mathcal{C}}(Y) + E_{\mathcal{A}}(S_{\mathcal{C}\parallel\mathcal{D}^*}(E_{\mathcal{C}}(Y)))\ \forall Y \in \langle\mathcal{B}\rangle. & (68)
\end{aligned}$$

It should be noted that we have not defined how to revise any $\mathcal{D}_k \in \mathcal{D}$ following the observation of $\mathcal{D}_j$. If we do not wish to observe $\mathcal{D}_j$ then we could treat the collection $\mathcal{D}_{j,P}$ as a member of $\mathcal{B}$. If however, having observed some $\mathcal{D}^* \subset \mathcal{D}$ we were to later observe $\mathcal{D}_k$ a sequential update of the form $[[\mathcal{B} \parallel \mathcal{D}^*]/[\mathcal{D}_{k,P} \parallel \mathcal{D}_j]]$ may not be the best way to utilise the specifications made in Definition 2 and will not equal $[\mathcal{B} \parallel \mathcal{D}^* \cup \mathcal{D}_k]$. To see this, consider $\mathcal{D}^* = \mathcal{D}_j$. Notice that whilst the belief structure $[\mathcal{B}_{k,P}]$ is specified, if $W \in \langle\mathcal{B}_{k,P}\rangle$, we cannot obtain $Cov(W, Y)$ for all $Y \in \langle\mathcal{D}_{j,P}\rangle$. In particular, we cannot deduce $Cov(\mathcal{B}^2_k, Y)$. Thus, revisions by $\mathcal{D}_j$ can only occur on the belief structure $[\mathcal{B}_k]$. Revision of $\mathcal{B}_k$ by $\mathcal{D}_k$, having first revised everything by $\mathcal{D}_j$ can only use the second-order structure of the relationship between $\mathcal{B}_k$ and $\mathcal{D}_k$. In contrast, equations (58) - (60) do utilise the full probabilistic between each $\mathcal{B}_j$ and $\mathcal{D}_j$.

8.1 Local computations for the revision

We commented that the revision provided by Definition 7 may be viewed as modifying the evidence into $[\mathcal{B}^*]$ and absorbing this using Bayes linear methods. Theorem 4 shows that this revision is consistent for separations on the augmented BLB graphical model. The modified evidence, $\{E(\mathcal{B}^* \parallel \mathcal{D}^*),\ S_{\mathcal{B}^*\parallel\mathcal{D}^*}(\mathcal{B}^*)\}$, is obtained, see Definition 6, by taking the evidence at each $\mathcal{B}_{(j)}$, $\{E(\mathcal{B}_{(j)} \parallel \mathcal{D}_{(j)}),\ S_{\mathcal{B}_{(j)}\parallel\mathcal{D}_{(j)}}(\mathcal{B}_{(j)})\}$, and absorbing it into $\mathcal{B}^*$ using a Bayes linear calculation. The net result of our calculations is that we compute the modified evidence, for each $j = 1, \ldots, s'$, $\{E(\mathcal{B}_{(j)} \parallel \mathcal{D}_{(j)}),\ S_{\mathcal{B}_{(j)}\parallel\mathcal{D}_{(j)}}(\mathcal{B}_{(j)})\}$ and absorb it into $\mathcal{B}$ using Bayes linear calculations. The consistency of Bayes linear adjustment and Theorems 4 and Theorems 5 of Goldstein & Wilkinson (2000) coupled with Definition 7 and Theorem 4 gives us the following theorem; equations (69) and (70) follow from equations (61) and (62) and equations (71) and (72) by inverting equations (69) and (70). Notice that if $(\mathcal{D}^*_P \perp\!\!\!\perp_S \mathcal{D}^\circ_P)|\mathcal{A}$ then $(\mathcal{B}^* \perp\!\!\!\perp_S \mathcal{B}^\circ)|\mathcal{A}$.

Theorem 5 *If $\mathcal{D}^\diamond$, $\mathcal{D}^\circ \subset \mathcal{D}$, $\mathcal{A} \subseteq \mathcal{B}$ are such that $(\mathcal{D}^\diamond_P \perp\!\!\!\perp_S \mathcal{D}^\circ_P)|\mathcal{A}$ then*

$$\begin{aligned}
&(i) \quad S_{\mathcal{A}\|\mathcal{D}^\diamond,\mathcal{D}^\circ}(\mathcal{A}) = (S^\dagger_{\mathcal{A}\|\mathcal{D}^\diamond}(\mathcal{A}) + S^\dagger_{\mathcal{A}\|\mathcal{D}^\circ}(\mathcal{A}) - I)^\dagger; && (69)\\
&(ii) \quad E^T(\mathcal{A} \parallel \mathcal{D}^\diamond, \mathcal{D}^\circ) = (E^T(\mathcal{A} \parallel \mathcal{D}^\diamond)S^\dagger_{\mathcal{A}\|\mathcal{D}^\diamond}(\mathcal{A}) + E^T(\mathcal{A} \parallel \mathcal{D}^\circ)S^\dagger_{\mathcal{A}\|\mathcal{D}^\circ}(\mathcal{A}) \\
&\qquad\qquad - E^T(\mathcal{A}))S_{\mathcal{A}\|\mathcal{D}^\diamond,\mathcal{D}^\circ}(\mathcal{A}); && (70)\\
&(iii) \quad S_{\mathcal{A}\|\mathcal{D}^\diamond} = (S_{\mathcal{A}\|\mathcal{D}^\diamond,\mathcal{D}^\circ}(\mathcal{A})^\dagger - S^\dagger_{\mathcal{A}\|\mathcal{D}^\circ}(\mathcal{A}) + I)^\dagger; && (71)\\
&(iv) \quad E^T(\mathcal{A} \parallel \mathcal{D}^\diamond) = (E^T(\mathcal{A} \parallel \mathcal{D}^\diamond, \mathcal{D}^\circ)S_{\mathcal{A}\|\mathcal{D}^\diamond,\mathcal{D}^\circ}(\mathcal{A})^\dagger \\
&\qquad\qquad - E^T(\mathcal{A} \parallel \mathcal{D}^\circ)S^\dagger_{\mathcal{A}\|\mathcal{D}^\circ}(\mathcal{A}) + E^T(\mathcal{A}))S_{\mathcal{A}\|\mathcal{D}^\diamond}. && (72)
\end{aligned}$$

Notice, in particular, that since $(\mathcal{D}^\diamond_P \perp\!\!\!\perp_S \mathcal{D}^\circ_P)|\mathcal{B}$, we may obtain the equivalent definitions to equations (63) and (64) that

$$E^T(\mathcal{B} \parallel \mathcal{D}^*) = (\sum_{j=1}^{s'} E^T(\mathcal{B} \parallel \mathcal{D}_{(j)})S^\dagger_{\mathcal{B}\|\mathcal{D}_{(j)}}(\mathcal{B}) - (s'-1)E^T(\mathcal{B}))S_{\mathcal{B}\|\mathcal{D}^*}(\mathcal{B}), \qquad (73)$$

where

$$S_{\mathcal{B}\|\mathcal{D}^*}(\mathcal{B}) = (\sum_{j=1}^{s'} S^\dagger_{\mathcal{B}\|\mathcal{D}_{(j)}}(\mathcal{B}) - (s-1)I)^\dagger(X). \qquad (74)$$

and so $Var(\mathcal{B} \parallel \mathcal{D}^*) = Var(\mathcal{B})S_{\mathcal{B}\|\mathcal{D}^*}(\mathcal{B})$.

9 Local computation for revision of the junction tree

Theorem 4 and Theorem 5 provide us with a method of calculating the belief revision using a local computation algorithm. The following algorithm is a straightforward modification of the local computation algorithm for global adjustment of the junction tree given in Section 3.4 of Goldstein & Wilkinson (2000). The algorithm is based on the junction tree for the moral graph of the augmented BLB graphical model of Definition 2. Jensen (1996), for example, provides details about the construction of the junction tree for the moral graph. Notice that, as a consequence of conditions 1. - 3. of Definition 2, each $\mathcal{D}_j$ appears in a single node on the junction tree, that node being $\{\mathcal{D}_j, \mathcal{B}_j\}$. We denote the nodes of the junction tree by $J_1, \ldots, J_u$, where $u \geq s+1$. We may label the nodes so that for $j = 1, \ldots, s$, $J_j = \{\mathcal{D}_j, \mathcal{B}_j\}$. Thus, if $t > s$ then $J_t \subseteq \mathcal{B}$. If J_r and J_t are connected on the junction tree with separator W_{rt} then we let $U_r = J_r \setminus W_{rt}$ and $U_t = J_t \setminus W_{rt}$, so that $(U_r \perp\!\!\!\perp_S U_t)|W_{rt}$. Notice that each $W_{rt} \subset \mathcal{B}$. At each node J_t, $t > s$, we store $E(J_t)$ and $Var(J_t)$. A current value for the revised belief transform $S(J_t) = S_{J_t\|\mathcal{D}^*}(J_t)$ and revised expectation $A(J_t) = E(J_t \parallel \mathcal{D}^*)$ are also stored, where initially $\mathcal{D}^* = \emptyset$ and so $S(J_t) = I$ and $A(J_t) = E(J_t)$. Later, $\mathcal{D}^* \subseteq \mathcal{D}$.

Theorems 4 and 5 provide the basis for a local computation algorithm for the revision of $\mathcal{B}$ by $\mathcal{D}$ as we now explain.

9.1 Entering evidence

For each $j = 1, \ldots, s$ at the node J_j we store $A(J_j) = E(\mathcal{B}_j \parallel \mathcal{D}_j)$ and $S(J_j) = S_{\mathcal{B}_j\|\mathcal{D}_j}(\mathcal{B}_j)$. Note that if J_t is connected to J_j then $W_{jt} = J_j \cap J_t = \mathcal{B}_j$.

9.2 Absorption

Suppose that the evidence $\{E(J_t \parallel \mathcal{D}_{(i)}),\ S_{J_t\parallel\mathcal{D}_{(i)}}(J_t)\}$ is obtained by node J_t, which has current revision information $A(J_t) = E(J_t \parallel \mathcal{D}^*)$ and $S(J_t) = S_{J_t\parallel\mathcal{D}^*}(J_t)$. Now $(\mathcal{D}_{(i),P} \perp\!\!\!\perp_S \mathcal{D}_{(i),P})|J_t$ and using equations (69) and (70), $S_{J_t\parallel\mathcal{D}^*,\mathcal{D}_{(i)}}(J_t)$ and $E(J_t \parallel \mathcal{D}^*, \mathcal{D}_{(i)})$ are calculated and these replace the old values of $S(J_t)$ and $A(J_t)$: the new evidence $\mathcal{D}_{(i)}$ is said to have been absorbed into the revision.

9.3 Message-passing

When requested for a message, a node J_t, for $t > s$, will compute $V(J_t) = Var(J_r)S(J_t)$ and then return the message $\{A(J_t),\ V(J_t)\}$ to the caller. For the nodes J_j, for $j \leq s$, we return the evidence message $\{A(J_j),\ S(J_j)\}$.

9.4 Processing a collect-phase message

The structure of the junction tree for the augmented BLB graphical model means that the only nodes that receive messages are the $J_{t'}$ for $t' > s$. If $J_{t'}$ receives the message $\{E(J_t \parallel \mathcal{D}^*),\ Var(J_t \parallel \mathcal{D}^*)\}$ from J_t, $t > s$, then $J_{t'}$ first extracts the marginals $E(W_{tt'} \parallel \mathcal{D}^*)$ and $Var(W_{tt'} \parallel \mathcal{D}^*)$ where $W_{tt'} = J_t \cap J_{t'}$ We then compute $S_{W_{tt'}\parallel\mathcal{D}^*}(W_{tt'}) = Var^{\dagger}(W_{tt'})Var(W_{tt'} \parallel \mathcal{D}^*)$. If $J_{t'}$ receives the message $\{A(J_t),\ S(J_t)\}$ from J_t, $t \leq s$, then $A(J_t) = E(W_{tt'} \parallel \mathcal{D}^*)$ and $S(J_t) = S_{W_{tt'}\parallel\mathcal{D}^*}(W_{tt'})$. On the collect-phase we have $(J_{t'} \perp\!\!\!\perp_S \mathcal{D}^*_P)|W_{tt'}$, so we use (65) and (68) to compute the message $\{E(J_{t'} \parallel \mathcal{D}^*),\ S_{J_{t'}\parallel\mathcal{D}^*}(J_{t'})\}$ ready for absorption by $J_{t'}$.

9.5 Processing a distribute-phase message

On the distribute-phase the message $\{E(J_t \parallel D^*, D^\circ),\ Var(J_t \parallel \mathcal{D}^*, \mathcal{D}^\circ)\}$ is received by the node $J_{t'}$ from the node J_t. $\mathcal{D}^\circ$ represents the information already absorbed by $J_{t'}$ and $\mathcal{D}^*$ the extra information to be absorbed into $J_{t'}$. $J_{t'}$ first extracts the marginals $E(W_{tt'} \parallel D^*, D^\circ)$ and $Var(W_{tt'} \parallel D^*, D^\circ)$ and then computes $S_{W_{tt'}\parallel D^*,D^\circ}(W_{tt'}) = Var^{\dagger}(W_{tt'})Var(W_{tt'} \parallel D^*, D^\circ)$. In the distribute phase, we have $(\mathcal{D}^* \perp\!\!\!\perp_S \mathcal{D}^\circ)|W_{tt'}$ and so utilise equations (71) and (72) to form the message $\{E(J_{t'} \parallel \mathcal{D}^*),\ S_{J_{t'}\parallel\mathcal{D}^*}(J_{t'})\}$ ready for absorption by $J_{t'}$.

9.6 Collection and distribution of evidence

9.6.1 Collecting evidence

From the $J_{t'}$, $t' > s$, pick an arbitrary root node and send it the message CollectEvidence. When a node, J_t receives this message, it sends the message to each of its other neighbours and processes and absorbs each message in turn. If $t \leq s$, the message $\{A(J_t),\ S(J_t)\}$ is returned to the caller, whilst if $t > s$, the message $\{A(J_t),\ V(J_t)\}$ is returned to the caller.

Upon completion of the collect phase, the revision at each node represents the revision by all nodes lower, with respect to the root node, than it in the junction tree. Notice that the nodes J_j for each $j = 1, \ldots, s$ have no node lower than them. We now delete each J_j from the junction tree and operate a distribute phase on the junction tree with nodes J_{s+1}, $\ldots$, J_u.

Tested	$E(\mathcal{M}(X_1) \parallel \mathcal{D})$	$E(\mathcal{M}(X_2) \parallel \mathcal{D})$	$E(\mathcal{M}(X_3) \parallel \mathcal{D})$	$E(\mathcal{M}(X_4) \parallel \mathcal{D})$	$E(\mathcal{M}(X_5) \parallel \mathcal{D})$
Prior	0.200000	0.200000	0.300000	0.400000	0.500000
I_1	0.190000	0.197500	0.296250	0.395000	0.493750
I_1, I_2	0.187757	0.187757	0.292654	0.390206	0.487757
$I_1 - I_3$	0.184047	0.184047	0.272238	0.385033	0.481291
$I_1 - I_4$	0.178454	0.178454	0.266899	0.353908	0.466241
$I_1 - I_5$	0.170551	0.170551	0.259356	0.338036	0.421556

Table 1: Revised expectations for single successful test in the specified subdomains.

9.6.2 Distributing evidence

The message DistributeEvidence is sent to the root node. The node J_t should, on receipt of this message, process and absorb any message, and to all of its other neighbours pass the message DistributeEvidence $\{A(J_t),\, V(J_t)\}$.

Upon completion of the distribute phase, each node on the junction tree has been revised by the total evidence, $\mathcal{D}$.

10 Partition testing example revisited

We return to the partition testing example of Section 6. A junction tree representation of the augmented BLB model consists of $k+1$ nodes: $\mathcal{M}(\mathcal{C})$ and the pairs $\{\mathcal{D}_g,\, \mathcal{M}(X_g)\}$ for each $g = 1, \ldots, k$. The k separators are the nodes $\mathcal{M}(X_g)$. We consider taking observations in various subdomains and revise our beliefs over the $\mathcal{M}(\mathcal{C})$. Theorems 4 and 5 show that we need only calculate the evidence messages $\{E(\mathcal{M}(\mathcal{X}_g) \parallel \mathcal{D}_g),\, S_{\mathcal{M}(X_g)\parallel\mathcal{D}_g}(\mathcal{M}(X_g))\}$. The assumption is made that each $\mathcal{M}(X_g)$ is Beta$(\alpha_g,\, \beta_g)$ and we follow the convention that we observe n_g tests in the I_gth subdomain and observe $x_g = 0$ incorrect outputs; that is, each observed test outcome is correct. Then, $E(\mathcal{M}(\mathcal{X}_g) \parallel \mathcal{D}_g) = (\alpha_g/(\alpha_g + beta_g + n_g))$ and

$$
\begin{aligned}
S_{\mathcal{M}(X_g)\parallel\mathcal{D}_g}(\mathcal{M}(X_g)) &= Var^{\dagger}(\mathcal{M}(X_g))Var(\mathcal{M}(X_g) \parallel \mathcal{D}_g) &\quad (75)\\
&= \frac{(\alpha_g+\beta_g)^2(\alpha_g+\beta_g+1)(\beta_g+n_g)}{(\alpha_g+\beta_g+n_g)^2(\alpha_g+\beta_g+n_g+1)\beta_g}, &\quad (76)
\end{aligned}
$$

and these evidence messages may be absorbed into $\mathcal{B}$.

We now consider the explicit example where our prior beliefs over $[\mathcal{M}(\mathcal{C})]$ are the same as in the first example of Coolen *et al.* (2001), see equation (40). We consider the effect where a single input is tested per subdomain. To show the comparisons between observations in different subdomains, we illustrate this by first considering a single input in the first subdomain, and then add in an additional observation in the second subdomain, then the third and so on.

The results of our analysis is shown in Table 1 and Table 2. Notice how, because of the positive correlations, observing successful tests reduces our uncertainty in all of the subdomains. We now consider comparing the revised quantities with the Bayes linear adjusted quantities. The results are shown in Table 3 and Table 4. Notice how each $VD(\mathcal{M}(X_g))$ is negative, the revised variances are all smaller than the analogous adjusted variances. As we pointed out in our analysis of the single subdomain observation, this is not surprisingly. Our ability to condition avoids us having to average the variance across all possible data points and the conditional variance for observing a test success is smaller than the Bayes linear version. This is carried through when we merge our observation. Contrast this with the

Tested	$V(\mathcal{M}(X_1) \parallel \mathcal{D})$	$V(\mathcal{M}(X_2) \parallel \mathcal{D})$	$V(\mathcal{M}(X_3) \parallel \mathcal{D})$	$V(\mathcal{M}(X_4) \parallel \mathcal{D})$	$V(\mathcal{M}(X_5) \parallel \mathcal{D})$
Prior	0.008000	0.008000	0.015000	0.020000	0.025000
I_1	0.007329	0.007958	0.014906	0.019832	0.024738
I_1, I_2	0.007293	0.007293	0.014815	0.019671	0.024486
$I_1 - I_3$	0.007242	0.007242	0.013250	0.019571	0.024329
$I_1 - I_4$	0.007175	0.007175	0.013190	0.017520	0.023850
$I_1 - I_5$	0.007104	0.007104	0.013125	0.017232	0.021564

Table 2: Revised variances for single successful test in the specified subdomains.

Tested	$ED(\mathcal{M}(X_1))$	$ED(\mathcal{M}(X_2))$	$ED(\mathcal{M}(X_3))$	$ED(\mathcal{M}(X_4))$	$ED(\mathcal{M}(X_5))$
I_1	0.000000	0.000000	0.000000	0.000000	0.000000
I_1, I_2	0.000103	0.000103	0.000062	0.000082	0.000103
$I_1 - I_3$	0.000317	0.000317	0.000401	0.000279	0.000349
$I_1 - I_4$	0.006644	0.006644	0.000806	0.000823	0.000810
$I_1 - I_5$	0.001125	0.001125	0.001348	0.001516	0.001309

Table 3: Differences between the revised expectations and the adjusted expectations for single successful test in the specified subdomains, $ED(\mathcal{M}(X_g)) = E(\mathcal{M}(X_g) \parallel \mathcal{D}) - E_{\mathcal{D}_P}(\mathcal{M}(X_g))$.

Tested	$VD(\mathcal{M}(X_1))$	$VD(\mathcal{M}(X_2))$	$VD(\mathcal{M}(X_3))$	$VD(\mathcal{M}(X_4))$	$VD(\mathcal{M}(X_5))$
I_1	-0.000271	-0.000017	-0.000038	-0.000068	-0.000106
I_1, I_2	-0.000284	-0.000284	-0.000074	-0.000131	-0.000205
$I_1 - I_3$	-0.000298	-0.000298	-0.000583	-0.000161	-0.000252
$I_1 - I_4$	-0.000307	-0.000307	-0.000589	-0.000588	-0.000341
$I_1 - I_5$	-0.000298	-0.000298	-0.000579	-0.000560	-0.000279

Table 4: Differences between the revised variances and the adjusted variances for single successful test in the specified subdomains, $VD(\mathcal{M}(X_g)) = Var(\mathcal{M}(X_g) \parallel \mathcal{D}) - Var_{\mathcal{D}_P}(\mathcal{M}(X_g))$.

observation that the revised means are all slightly larger than the corresponding adjusted means.

Acknowledgements

This work has been supported by grant GR\M76775 from EPSRC.

References

Bachman, G. and L. Narici (2000). *Functional Analysis.* Mineola, New York: Dover.

Cambanis, S. (1982). Conditional probability and expectation. In S. Kotz, N. L. Johnson, and C. B. Read (Eds.), *Encyclopedia of Statistical Sciences*, Volume 2, pp. 109–119. New York: Wiley.

Coolen, F. P. A., M. Goldstein, and M. Munro (2001). Generalized partition testing via Bayes linear methods. *Information and Software Technology 43*, 783–793.

Cowell, R. G., A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter (1999). *Probabilistic Networks and Expert Systems.* New York: Springer.

Dawid, A. P. (1979). Conditional independence in statistical theory (with discussion). *J. R. Statist. Soc. B 41*, 1–31.

Dawid, A. P. (1980). Conditional independence for statistical operations. *Ann. Statist. 8*, 598–617.

de Finetti, B. (1974). *Theory of Probability*, Volume 1. London: Wiley.

de Finetti, B. (1975). *Theory of Probability*, Volume 2. London: Wiley.

Goldstein, M. (1979). The variance modified linear Bayes estimator. *J.R.Statist. Soc. B 41*, 96–100.

Goldstein, M. (1981). Revising previsions: a geometric interpretation (with discussion). *J. R. Statist. Soc. B 43*, 105–130.

Goldstein, M. (1983). General variance modifications for linear Bayes estimators. *J. Amer. Statist. Ass. 78*, 616–618.

Goldstein, M. (1986). Exchangeable belief structures. *J. Amer. Statist. Ass. 81*, 971–976.

Goldstein, M. (1990). Influence and Belief Adjustment. In R. M. Oliver and J. Q. Smith (Eds.), *Influence Diagrams, Belief Nets and Decision Analysis*, pp. 143–174. Wiley.

Goldstein, M. (1991). Belief transforms and the comparison of hypotheses. *Ann. Statist. 19*, 2067–2089.

Goldstein, M. (1994). Revising exchangeable beliefs: subjectivist foundations for the inductive argument. In P. Freeman and A. Smith (Eds.), *Aspects of Uncertainty: A Tribute to D.V.Lindley*, pp. 201–222. Wiley.

Goldstein, M. (1999). Bayes linear analysis. In S. Kotz, C. B. Read, and D. L. Banks (Eds.), *Encyclopedia of Statistical Sciences Update Volume 3*, pp. 29–34. Chichester: Wiley.

Goldstein, M. and D. J. Wilkinson (2000). Bayes linear analysis for graphical models: The geometric approach to local computation and interpretive graphics. *Statistics and Computing 10*, 311–324.

Jensen, F. V. (1996). *An introduction to Bayesian networks.* London: University College London Press.

Lauritzen, S. L. (1996). *Graphical Models.* Oxford: Oxford Science Publications.

Lauritzen, S. L., A. P. Dawid, B. N. Larsen, and H.-G. Leimer (1990). Independence properties of directed Markov fields. *Networks 20*, 491–505.

Pearl, J. (1988). *Probabilistic Inference in Intelligent Systems.* San Mateo, California: Morgan Kaufman.

Rees, K., F. P. A. Coolen, M. Goldstein, and D. A. Wooff (2001). Managing the uncertainties of software testing: a Bayesian approach. *Quality and Reliability Engineering International 17*, 191–203.

Smith, J. Q. (1989). Influence diagrams foe statistical modelling. *Annals of Statistics 17*, 654–672.

Weyuker, E. J. and T. J. Ostrand (1980). Theories of program testing and the application of revealing subdomains. *IEEE Transactions on Software Engineering 6*, 236–246.

Appendix

Proof of Lemma 1 - We consider separations on the relevant moral graphs, see Lauritzen *et al.* (1990), constructed from the augmented BLB graphical model where the node $\mathcal{D}_j$ represents the collection of random quantities $\mathcal{D}_{j,P}$. For three collections of nodes of interest, $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, we denote the moral graph constructed over these nodes as $\mathcal{G}_M(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$. $\mathcal{G}_M(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$ is constructed by restricting $\mathcal{G}$ to the nodes $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$ and all their ancestors and all the arcs between the nodes in this restricted collection. We draw an arc between any two nodes that share a child but are not currently joined and then drop all arrows to form $\mathcal{G}_M(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$. If the nodes of $\mathcal{G}_M(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$ are the same as $\mathcal{G}$ then we tern $\mathcal{G}_M(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$ the full moral graph, written $\mathcal{G}_M$.

The first statement is trivial by the observation that since $pa(\mathcal{D}_j) = \mathcal{B}_j$ and $ch(\mathcal{D}_j) = \emptyset$, the moral graph $\mathcal{G}_M(\mathcal{A}, \mathcal{D}_j, \mathcal{C})$ is the moral graph $\mathcal{G}_M(\mathcal{A}, \mathcal{B}_j, \mathcal{C})$ with the addition of the node $\mathcal{D}_j$ and an arc between $\mathcal{B}_j$ and $\mathcal{D}_j$.

The second statement is trivial if $\mathcal{B}_j \in \mathcal{C}$ and so we consider the case where $\mathcal{B}_j$ is not a member of $\mathcal{C}$. If neither $(\mathcal{A} \perp\!\!\!\perp_S \mathcal{D}_{j,P})|\mathcal{C}$ nor $(\mathcal{F} \perp\!\!\!\perp_S \mathcal{D}_{j,P})|\mathcal{C}$ then there is a node $\mathcal{A}_{\mathcal{D}_j} \in \mathcal{A}$ such that there is a path on $\mathcal{G}_M(\mathcal{A}, \mathcal{D}_j, \mathcal{C})$ between $\mathcal{A}_{\mathcal{D}_j}$ and $\mathcal{D}_j$ which does not intersect $\mathcal{C}$ and there is a node $\mathcal{F}_{\mathcal{D}_j} \in \mathcal{F}$ such that there is a path on $\mathcal{G}_M(\mathcal{F}, \mathcal{D}_j, \mathcal{C})$ between $\mathcal{F}_{\mathcal{D}_j}$ and $\mathcal{D}_j$ which does not intersect $\mathcal{C}$. Since any path to $\mathcal{D}_j$ must pass through $\mathcal{B}_j$, there is a path $\mathcal{A}_{\mathcal{D}_j}$ to $\mathcal{B}_j$ on $\mathcal{G}_M(\mathcal{A}, \mathcal{D}_j, \mathcal{C})$ which does not intersect $\mathcal{C} \cup \mathcal{D}_j$ and there is a path $\mathcal{F}_{\mathcal{D}_j}$ to $\mathcal{B}_j$ on $\mathcal{G}_M(\mathcal{F}, \mathcal{D}_j, \mathcal{C})$ which does not intersect $\mathcal{C} \cup \mathcal{D}_j$. Both of these paths are available on the moral graph $\mathcal{G}_M(\mathcal{A}, \mathcal{F}, \mathcal{C} \cup \mathcal{D}_j)$ and so we can combine these two paths to show there is a path between $\mathcal{A}_{\mathcal{D}_j}$ and $\mathcal{F}_{\mathcal{D}_j}$ on $\mathcal{G}_M(\mathcal{A}, \mathcal{F}, \mathcal{C} \cup \mathcal{D}_j)$ which does not intersect $\mathcal{C} \cup \mathcal{D}_{j,P}$. This is a contradiction to the statement that $(\mathcal{A} \perp\!\!\!\perp_S \mathcal{F})|\mathcal{C} \cup \mathcal{D}_{j,P}$ and so the second statement follows. □

Proof of Theorem 1 - From the first statement of Lemma 1, we have that $(\mathcal{A} \perp\!\!\!\perp_S \mathcal{D}_{j,P})|\mathcal{C}$ $\Rightarrow (\mathcal{A} \perp\!\!\!\perp_S \mathcal{B}_j)|\mathcal{C}$. Now,

$$\begin{aligned} E(\mathcal{A} \parallel \mathcal{D}_j) &= E(E_{\mathcal{B}_j}(\mathcal{A})|\mathcal{D}_j) && (77)\\ &= E(E_{\mathcal{B}_j}(E_{\mathcal{C}}(\mathcal{A}))|\mathcal{D}_j) && (78)\\ &= E(E_{\mathcal{C}}(\mathcal{A}) \parallel \mathcal{D}_j), && (79) \end{aligned}$$

where equation (78) follows from equation (16) of Goldstein & Wilkinson (2000) since $(\mathcal{A}\perp\!\!\!\perp_S \mathcal{B}_j)|\mathcal{C}$.

$$\begin{aligned} Var(\mathcal{A} \parallel \mathcal{D}_j) &= Var_{\mathcal{B}_j}(\mathcal{A}) + Var(E_{\mathcal{B}_j}(\mathcal{A})|\mathcal{D}_j) && (80)\\ &= Var_{\mathcal{C}}(\mathcal{A}) + Var_{\mathcal{B}_j}(E_{\mathcal{C}}(\mathcal{A})) + Var(E_{\mathcal{B}_j}(E_{\mathcal{C}}(\mathcal{A}))|\mathcal{D}_j) && (81)\\ &= Var_{\mathcal{C}}(\mathcal{A}) + Var(E_{\mathcal{C}}(\mathcal{A}) \parallel \mathcal{D}_j), && (82)\end{aligned}$$

where equation (81) follows from equations (16) and (17) of Goldstein & Wilkinson (2000) since $(\mathcal{A}\perp\!\!\!\perp_S \mathcal{B}_j)|\mathcal{C}$.

$$\begin{aligned} Cov(\mathcal{C}, \mathcal{A} \parallel \mathcal{D}_j) &= Cov_{\mathcal{B}_j}(\mathcal{C}, \mathcal{A}) + Cov(E_{\mathcal{B}_j}(\mathcal{C}), E_{\mathcal{B}_j}(\mathcal{A}) \parallel \mathcal{D}_j) && (83)\\ &= Cov_{\mathcal{B}_j}(\mathcal{C}, E_{\mathcal{C}}(\mathcal{A})) + Cov(E_{\mathcal{B}_j}(\mathcal{C}), E_{\mathcal{B}_j}(E_{\mathcal{C}}(\mathcal{A}))|\mathcal{D}_j) && (84)\\ &= Cov(\mathcal{C}, E_{\mathcal{C}}(\mathcal{A}) \parallel \mathcal{D}_j), && (85)\end{aligned}$$

where equation (84) follows from equations (16) and (18) of Goldstein & Wilkinson (2000) since $(\mathcal{A}\perp\!\!\!\perp_S \mathcal{B}_j)|\mathcal{C}$. □

Proof of Theorem 2 - If $(\mathcal{A}\perp\!\!\!\perp_S \mathcal{F})|\mathcal{C}\cup\mathcal{D}_{j,P}$ then we may assume, without loss of generality, that $(\mathcal{A}\perp\!\!\!\perp_S \mathcal{D}_{j,P})|\mathcal{C}$; see the second statement of Lemma 1. Applying equation (28), we have

$$\begin{aligned} Cov(\mathcal{A}, \mathcal{C} \parallel \mathcal{D}_j) &= Cov(E_{\mathcal{C}}(\mathcal{A}), \mathcal{C} \parallel \mathcal{D}_j) && (86)\\ &= Cov(\mathcal{A}, \mathcal{C})Var^{\dagger}(\mathcal{C})Var(\mathcal{C} \parallel \mathcal{D}_j). && (87)\end{aligned}$$

Thus,

$$\begin{aligned} &Cov(\mathcal{A}, \mathcal{C} \parallel \mathcal{D}_j)Var^{\dagger}(\mathcal{C} \parallel \mathcal{D}_j)Cov(\mathcal{C}, \mathcal{F} \parallel \mathcal{D}_j) = \\ &\quad Cov(\mathcal{A}, \mathcal{C})Var^{\dagger}(\mathcal{C})Cov(\mathcal{C}, \mathcal{F} \parallel \mathcal{D}_j) && (88)\\ &= Cov(\mathcal{A}, \mathcal{C})Var^{\dagger}(\mathcal{C})\{Cov_{\mathcal{B}_j}(\mathcal{C}, \mathcal{F}) + Cov(E_{\mathcal{B}_j}(\mathcal{C}), E_{\mathcal{B}_j}(\mathcal{F})|\mathcal{D}_j)\} && (89)\\ &= Cov(\mathcal{A}, \mathcal{C})Var^{\dagger}(\mathcal{C})\{Cov(\mathcal{C}, \mathcal{B}_j)Var^{\dagger}(\mathcal{B}_j)(Cov(\mathcal{B}_j, E_{\mathcal{B}_j}(\mathcal{F})|\mathcal{D}_j) - Cov(\mathcal{B}_j, \mathcal{F})) + \\ &\quad Cov(\mathcal{C}, \mathcal{F})\}. && (90)\end{aligned}$$

Now, since $(\mathcal{A}\perp\!\!\!\perp_S \mathcal{F})|\mathcal{C}$ then we have $Cov(\mathcal{A}, \mathcal{C})Var^{\dagger}(\mathcal{C})Cov(\mathcal{C}, \mathcal{F}) = Cov(\mathcal{A}, \mathcal{F})$. Additionally, since $(\mathcal{A}\perp\!\!\!\perp_S \mathcal{D}_{j,P})|\mathcal{C}$ then, from the first statement of Lemma 1, we have $(\mathcal{A}\perp\!\!\!\perp_S \mathcal{B}_j)|\mathcal{C}$ and thus $Cov(\mathcal{A}, \mathcal{C})Var^{\dagger}(\mathcal{C})Cov(\mathcal{C}, \mathcal{B}_j) = Cov(\mathcal{A}, \mathcal{B}_j)$. Substituting these two into equation (90) and rearranging gives

$$\begin{aligned} &Cov(\mathcal{A}, \mathcal{C} \parallel \mathcal{D}_j)Var^{\dagger}(\mathcal{C} \parallel \mathcal{D}_j)Cov(\mathcal{C}, \mathcal{F} \parallel \mathcal{D}_j) = \\ &\quad Cov_{\mathcal{B}_j}(\mathcal{A}, \mathcal{F}) + Cov(\mathcal{A}, \mathcal{B}_j)Var^{\dagger}(\mathcal{B}_j)Cov(\mathcal{B}_j, E_{\mathcal{B}_j}(\mathcal{F})|\mathcal{D}_j) && (91)\\ &= Cov_{\mathcal{B}_j}(\mathcal{A}, \mathcal{F}) + Cov(E_{\mathcal{B}_j}(\mathcal{A}), E_{\mathcal{B}_j}(\mathcal{F})|\mathcal{D}_j) && (92)\\ &= Cov(\mathcal{A}, \mathcal{F} \parallel \mathcal{D}_j), && (93)\end{aligned}$$

and hence the result. □