Bayesian Analysis (2014) **9**, Number 1, pp. 77–96

Bayes Linear Sufficiency in Non-exchangeable Multivariate Multiple Regressions

David A. Wooff [*]

Abstract. We consider sufficiency for Bayes linear revision for multivariate multiple regression problems, and in particular where we have a sequence of multivariate observations at different matrix design points, but with common parameter vector. Such sequences are not usually exchangeable. However, we show that there is a sequence of transformed observations which is exchangeable and we demonstrate that their mean is sufficient both for Bayes linear revision of the parameter vector and for prediction of future observations. We link these ideas to making revisions of belief over replicated structure such as graphical templates of model relationships. We show that the sufficiencies lead to natural residual collections and thence to sequential diagnostic assessments. We show how each finite regression problem corresponds to a parallel implied infinite exchangeable sequence which may be exploited to solve the sample-size design problem. Bayes linear methods are based on limited specifications of belief, usually means, variances, and covariances. As such, the methodology is well suited to high-dimensional regression problems where a full Bayesian analysis is difficult or impossible, but where a linear Bayes approach offers a pragmatic way to combine judgements with data in order to produce posterior summaries.

Keywords: Bayes linear, Sufficient, Multivariate multiple regression, Approximate Bayesian, Residual space, Diagnostics, Sequential, Sample-size design.

1 Introduction

We consider sufficiency for Bayes linear (BL) revision for multivariate multiple regression problems of the form $Y = X\beta + \epsilon$, where Y is a response vector, X is a design matrix, β is a vector of parameters, and ϵ is a vector of error quantities. In particular, we suppose that we will take n vector observations $Y_1, \ldots, Y_n$ at points $X_1, \ldots, X_n$ with error vectors $\epsilon_1, \ldots, \epsilon_n$, where each design point X_i is a matrix. The aim is to revise beliefs over β. This is a familiar problem which has been extensively addressed from classical and traditional Bayesian perspectives. What has not been addressed is the role of sufficiency in the multivariate BL setting, largely because the sequence of vector observables is not exchangeable. The aim of this paper is to reveal that there are sufficiencies which may be exploited in this context. These sufficiencies vastly simplify the computations required for the kinds of high-dimensional problems for which we use BL methods.

Our perspective is twofold. First, BL methodology, based on expectation rather than probability as a primitive, offers a rich pragmatic alternative to traditional Bayesian

[*] Durham University. Email: d.a.wooff@durham.ac.uk

DOI:10.1214/13-BA847

methods. Secondly, BL methods offer a practical linear approximation to full Bayesian methods for very high dimensional problems where the burden of full multi-dimensional probabilistic specification is too heavy or impossible, where computational tractability is important, and in particular where there is a need to prioritize for diagnostic insights. Recent examples motivated by such considerations include analysis of computer simulators (Cumming and Goldstein 2009), climate prediction (Goldstein and Rougier 2006, 2009; Williamson et al. 2012), galaxy formation (Vernon et al. 2010), and other careful high-dimensional modelling (Rougier and Kern 2010; Gosling et al. 2013). Because this material serves two audiences, in this account we will tend to mix terminology in order to make the content more familiar to those more used to traditional Bayesian approaches, so for example we blur the distinction between adjustment and revision, with context depending on one's choice of paradigm.

Section 2 outlines the basic concepts and calculations employed in BL methodology, introduces the idea of BL sufficiency, and identifies sufficiency with separation within graphical models for such regession problems. In Section 3 we introduce a motivating example. In Section 4 we establish notation and lemmas for subsequent theory. In Section 5 we derive sufficiency results for multiple multivariate regression parameters. In Section 6 we extend these results for prediction. In Section 7 we discuss the implications of sufficiency for simplifying residual analyses and show a brief example of such analysis. In Section 8 we show how the theory extends to problems where the parameter set β varies according to design matrix. In Section 9 we exploit the sufficiency construction to show how we may calculate approximately the sample size necessary to achieve specified variance reductions for any linear combination of parameters.

2 Bayes linear methods: an outline

BL methods arise either by taking expectation as the primitive quantification of degree of belief, replacing probability (de Finetti 1937, 1974; Goldstein 1981), or through simple linear approximations to full Bayesian analyses. They require only limited prior judgements and are easy to calculate, and so are well-suited to otherwise intractable problems. Basic definitions and calculations for the BL approach are briefly as follows. Further details may be found in Goldstein and Wooff (2007). The *adjusted expectation* (informally, the posterior mean) for collection B given collection D is

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

When we observe the collection D as $D = d$, we may evaluate (1) by replacing $[D-\mathrm{E}(D)]$ by $[d - \mathrm{E}(D)]$. We partition the variance matrix of B into components $\mathrm{Var}(B) = \mathrm{Var}(\mathrm{E}_D(B)) + \mathrm{Var}_D(B)$, being respectively the *resolved variance matrix* and the *adjusted variance matrix* (i.e. explained and residual variation), the residual portion being computed as

$$\mathrm{Var}_D(B) = \mathrm{Var}(B) - \mathrm{Cov}(B, D)\mathrm{Var}(D)^{\dagger}\mathrm{Cov}(D, B). \tag{2}$$

$A^{\dagger}$ is the Moore-Penrose generalized inverse, and we restrict attention to non-negative definite variance matrices with bounded positive trace. $\mathrm{E}_D(B)$ and $\mathrm{Var}_D(B)$ correspond informally with the conditional mean and variance $\mathrm{E}(B|D)$ and $\mathrm{Var}(B|D)$ in traditional settings.

A measure of the relative difference between the data d and their prior expectations $\mathrm{E}(D)$, is the *discrepancy*, $\mathrm{Dis}(d)$, the Mahalanobis distance between d and $\mathrm{E}(D)$:

$$\mathrm{Dis}(d) = [d - \mathrm{E}(D)]^T \mathrm{Var}(D)^\dagger [d - \mathrm{E}(D)].$$

A priori, $\mathrm{E}(\mathrm{Dis}(D)) = \mathbf{rank}\{\mathrm{Var}(D)\}$, so that the *discrepancy ratio*

$$\mathrm{Dr}(d) = \mathrm{Dis}(d)/\mathbf{rank}\{\mathrm{Var}(D)\},$$

is a standardized measure for the diagnostic: $\mathrm{E}(\mathrm{Dr}(D)) = 1$. Large changes in expectation coupled to small portions of variance explained would be quite surprising. Small changes in expectation coupled to large changes in variance would also be surprising, albeit in a different way. We derive similar diagnostics for observed adjusted expectations; the *adjustment discrepancy* is:

$$\mathrm{Dis}_d(B) = [\mathrm{E}_d(B) - \mathrm{E}(B)]^T [\mathrm{Cov}(D, B)\mathrm{Var}(D)^\dagger \mathrm{Cov}(D, B)]^\dagger [\mathrm{E}_d(B) - \mathrm{E}(B)].$$

This is the squared change in expectation from prior to posterior, relative to variance explained, and can be compared to its expected value:

$$\mathrm{E}(\mathrm{Dis}_d(B)) = r_{\mathbb{T}} = \mathbf{rank}\{\mathrm{Cov}(D, B)\}.$$

The *resolution transform matrix* is defined as

$$\mathbb{T}_{B:D} = \mathrm{Var}(B)^\dagger \mathrm{Cov}(B, D)\mathrm{Var}(D)^\dagger \mathrm{Cov}(D, B).$$

It is important because its eigenstructure provides the canonical form for a problem. The canonical directions $G_1, \ldots, G_{r_B}$, where $r_B = \mathbf{rank}\{\mathrm{Var}(B)\}$, derive from the normed right eigenvectors of $\mathbb{T}_{B:D}$, which we write $\tilde{v}_1, \ldots, \tilde{v}_{r_B}$, ordered by eigenvalues $1 \geq \lambda_1 \geq \ldots \geq \lambda_{r_B} \geq 0$, and scaled, for each i, as $\tilde{v}_i^T \mathrm{Var}(B)\tilde{v}_i = 1$. The canonical quantities are then defined as $G_i = \tilde{v}_i^T (B - \mathrm{E}(B))$. Properties of the canonical representation include $\mathrm{E}(G_i) = 0$, $\mathrm{Var}(G_i) = 1$, $\mathrm{Cov}(G_i, G_j) = 0$, and $\mathrm{Var}_D(G_i) = 1 - \lambda_i$. The eigenvalues are thus resolved variances for these uncorrelated components. The canonical form is useful (1) for revealing the structural implications of belief specifications, for example revealing those linear combinations for which data are expected to be informative/uninformative; (2) for attaching appropriate diagnostics to uncorrelated components of the model; (3) for sample-size design.

2.1 Bayes linear sufficiency

Consider a *second-order exchangeable* sample of vector random quantities $D_{(n)} = D_1, \ldots, D_n$ which respects exchangeability with another vector of random quantities B. That is, $\mathrm{E}(D_i)$ and $\mathrm{Var}(D_i)$ are the same for all i, $\mathrm{Cov}(D_i, D_j)$ is the same for all $i \neq j$, and $\mathrm{Cov}(D_i, B)$ is the same for all i. Then the sample mean vector $\bar{D} = \frac{1}{n} \sum D_i$

is BL *sufficient* for $D_{(n)}$ for adjusting B. Alternatively, this is the notion of sufficiency which we may apply when we constrain ourselves to second-order specifications within the full Bayesian setting. The practical implication, as for sufficiency in the classical and full Bayesian paradigms, is to simplify the computations required for posterior analysis. We derive additional benefits in the BL setting, for example in making diagnostic assessments. In particular, the canonical form for a sample size n can be constructed from the canonical form for a sample size $n = 1$, using an underlying eigenvalue relationship. This permits us to quantify for design purposes the implications of sample size changes on uncertainties for the variables of interest. These ideas are developed fully in Goldstein and Wooff (1998, 2007). Note that the identification of low-dimensional sufficient summaries is key to approximate Bayesian computation (Blum et al. 2013).

2.2 Sufficiency as separation between plates in a graphical model

Every multivariate multiple regression problem corresponds to a BL graphical model (Goldstein and Wooff 2007) which helps to clarify the role of sufficiency. For such graphs, the fundamental notion is of BL *separation* (Goldstein 1986), for which the notation $\lfloor A \perp\!\!\!\perp B \rfloor \,/\, C$ indicates that collections A, B are separated by a collection C. Separation on the directed acyclic graph is the property that for collections (nodes) A, B, C, we have $\mathrm{E}_{C\cup A}(B) = \mathrm{E}_C(B)$; $\mathrm{Var}_{C\cup A}(B) = \mathrm{Var}_C(B)$; and $\mathbb{T}_{B:C\cup A} = \mathbb{T}_{B:C}$. That is, C is BL sufficient for A for adjusting B. If we want to revise beliefs for B knowing C and A, we can discard A. One important consequence of belief separation is that covariances between separated structures may be evaluated via the separator, as follows.

Lemma 1.

$$\lfloor A \perp\!\!\!\perp B \rfloor \,/\, C \Longleftrightarrow \mathrm{Cov}(A, B) = \mathrm{Cov}(A, C)\mathrm{Var}(C)^{\dagger}\mathrm{Cov}(C, B).$$

Belief separation is a generalized conditional independence property (Goldstein 1990). BL graphical models form the BL analogue of Bayesian belief networks, with similar rules for node and arc operations, construction of junction trees and propagation of information (Goldstein and Wilkinson 2000; Goldstein and Wooff 2007; Wilkinson 1998).

3 Example: multivariate regression with correlated errors

This example, adapted from Box and Tiao (1973), is discussed in Goldstein and Wooff (2007). A chemical process leads to a product U and a by-product V. The yields of the products are thought to be related to the temperature of the process, $\tilde{X}$. Twelve experiments are performed with different temperature settings to study the effect of temperature. In performing the analysis, we transform the temperature measurements to $X = (\tilde{X} - 177.86)/100$, where 177.86 is the mean temperature in degrees Fahrenheit. The data (Table 1) are plotted in Figure 1. The model suggested to explain relationships between the quantities is as follows:

$$U_i = a + bx_i + e_i, \qquad \text{and} \qquad V_i = c + dx_i + f_i, \quad i = 1, \ldots, 12. \tag{3}$$

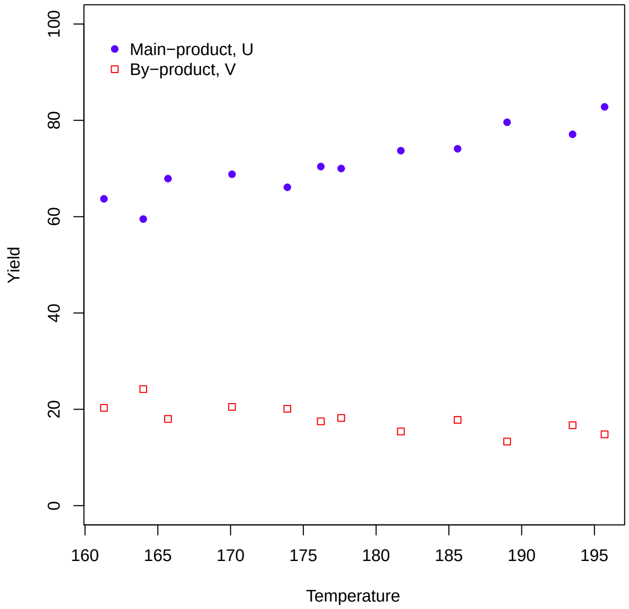

Figure 1: Yield of main-product U and by-product V at twelve temperatures $\tilde{X}$.

The model reflects the beliefs that the relationships between yields U, V and temperature X are approximately linear in X over the given range of temperature values. The intercept terms a, c indicate the yields for average temperature settings, whilst the slopes of the regressions are given by b, d. The models incorporate error components e_i, f_i. Separate runs of the experiment are independent; however, in any particular run it is felt that the error components will be correlated because slight aberrations in reaction conditions or analytical procedures could simultaneously affect both product yields. We will thus suppose that $e_1, e_2, \ldots$ are an uncorrelated sequence of error components with expectation zero and variance σ_e^2; that $f_1, f_2, \ldots$ are an uncorrelated sequence of error components with expectation zero and variance σ_f^2; and that all pairs of error components e_i, f_j are uncorrelated except for $\text{Cov}(e_i, f_i) = \sigma_{ef}$. Prior beliefs over these quantities were specified as follows. For the error quantities, $\sigma_e^2 = 6.25$, $\sigma_f^2 = 4$, $\sigma_{ef} = 2.5$, so that the correlation between the two error components for any given run is about 0.5. We specified (for details see Goldstein and Wooff (2007)) the following prior expectations and covariances between the regression coefficients:

$$\text{E}(\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}) = \begin{bmatrix} 75 \\ 40 \\ 20 \\ -30 \end{bmatrix}, \quad \text{Var}(\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}) = \begin{bmatrix} 4 & -6 & -1 & 0 \\ -6 & 225 & 0 & -90 \\ -1 & 0 & 1 & -2.4 \\ 0 & -90 & -2.4 & 144 \end{bmatrix}. \tag{4}$$

i	1	2	3	4	5	6	7	8	9	10	11	12
$\tilde{X}_i$	161.3	164.0	165.7	170.1	173.9	176.2	177.6	181.7	185.6	189.0	193.5	195.7
U_i	63.7	59.5	67.9	68.8	66.1	70.4	70.0	73.7	74.1	79.6	77.1	82.8
V_i	20.3	24.2	18.0	20.5	20.1	17.5	18.2	15.4	17.8	13.3	16.7	14.8

Table 1: Yield of main-product U and by-product V at twelve temperatures $\tilde{X}$.

Our focus is on revising beliefs over the regression coefficients, which we arrange as the collection $\beta = \{a, b, c, d\}$. We arrange the yields as the collections $U = (U_1 \ \ldots \ U_{12})$ and $V = (V_1 \ \ldots \ V_{12})$, and the sequence of observables as $Y_i = \{U_i, V_i\},\ i = 1, 2, \ldots, 12$. The error vectors are $\epsilon_i = \{e_i, f_i\}$, i=1,2,...,12. The design matrices are

$$X_i = \begin{bmatrix} 1 & x_i & 0 & 0 \\ 0 & 0 & 1 & x_i \end{bmatrix},\ i = 1, 2, \ldots, 12,$$

so that

$$Y_i = X_i\beta + \epsilon_i, \ \ i = 1, 2, \ldots, 12.$$

Note that the sequence of vectors $Y_1, Y_2, \ldots$ is not exchangeable because the design matrices $X_1, X_2, \ldots$ typically differ. As such we cannot employ BL sufficiency directly. Thus, the focus of what follows is to show that there is a transformation of the multivariate multiple regression problem which does lead to BL sufficiency of the sample mean of transformed observables, and so which leads to virtually all the exploitable qualities offered by exchangeable sequences.

3.1 Graphical representation using plates

A graphical model (more properly, a BL influence diagram) for this example is shown in Figure 2. This is adapted from Figure 10.2 of Goldstein and Wooff (2007). The graph is iteratively constructed from a consistent ordering of nodes according to rules given in Goldstein and Wooff (2007, section 10.5). Directions of arrows reflect the order of nodes in that consistent ordering: for convenience we have chosen an alphabetic ordering where allowable. Different graphs can result from a different choice of ordering. The arc directions have an implication for construction of the corresponding junction tree and thence to belief propagation and sequential local computation.

A general reduced form of the graph is shown in Figure 3. The i-subscripted quantities, $K_i = \{U_i, e_i, f_i, V_i, x_i\}$, are separated from the j-subscripted quantities,

$K_j = \{U_j, e_j, f_j, V_j, x_j\}$, by the subcollection of parameters, $\beta = \{a, b, c, d\}$. That is, $\lfloor K_i \perp\!\!\!\perp K_j \rfloor \ / \ \beta$. Moreover, K_i has the same internal structure as K_j, and the arcs between K_i and β are the same as those between K_j and β. Such duplication is typical when random quantities are constructed to represent error terms and observables which are connected through an underlying model. We indicate such duplicated structure on the graph as a *plate*: we include a single collection of nodes K_i on the graph, draw a dashed line about the collection, and indicate how many times this plate is repeated. A similar plate representation exists for every linear model of the kind considered in this paper. Correspondingly, a problem whose BL graphical model is representable in such plate form may be analysed via the sufficiencies which we identify in later sections.

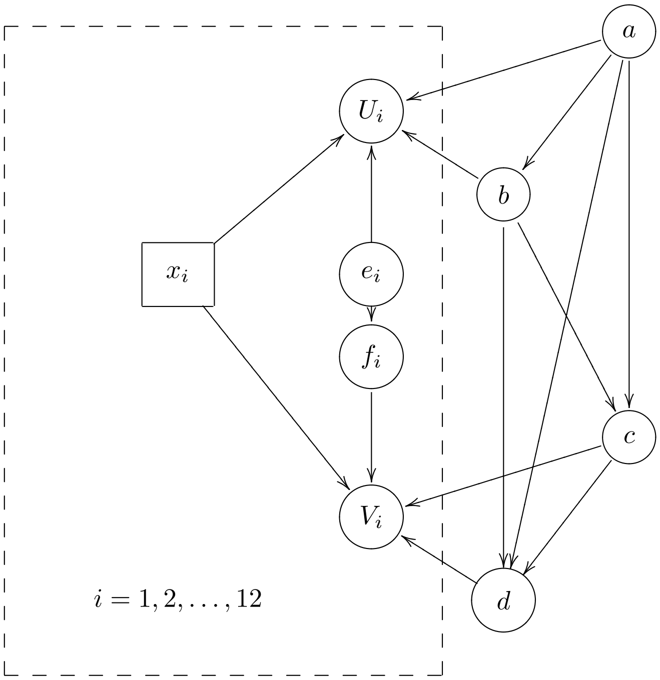

Figure 2: BL graphical model with a plate indicating repeated structure.

4 Notation and covariance structures

In this section we introduce some notation and summarise the expectation and covariance structures we will need in deriving subsequent theorems on sufficiency. We will suppose that Y_i is a $k \times 1$ vector of observables, β is an $m \times 1$ vector of parameters, X_i is a $k \times m$ design matrix, and ϵ_i is a $k \times 1$ vector of error quantities. Gather the vectors Y_i into the $nk \times 1$ vector $Y^T = [Y_1^T \ \ldots \ Y_n^T]$, the vectors ϵ_i likewise into the $nk \times 1$ vector ϵ, and stack the design matrices $X_1, \ldots, X_n$ as the $nk \times m$ matrix $X^T = [X_1^T \ \ldots X_n^T]$. We thus have $Y = X\beta + \epsilon$. We will assume the following basic prior specification. For

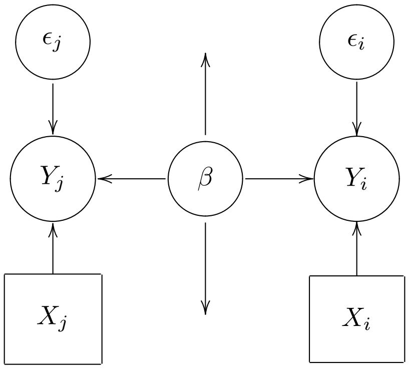

Figure 3: A reduced template form for Figure 2.

$i, j = 1, \ldots, n$ and $i \neq j$,

$$\text{Var}(\beta) = \Gamma, \;\; \text{Var}(\epsilon_i) = W_i, \;\; \text{Cov}(\epsilon_i, \epsilon_j) = 0, \;\; \text{Cov}(\beta, \epsilon_i) = 0. \tag{5}$$

We will assume without loss of generality that Γ is a positive definite $m \times m$ matrix; otherwise we can transform the elements of β to an equivalent set of full rank components. We will require $W_1, \ldots, W_n$ to be positive definite $k \times k$ variance matrices. $\text{E}(\beta)$ is as specified a priori, and we assume that $\text{E}(\epsilon_i) = 0$ for all i. For reasons which will become clear, we introduce transformations of the observables as

$$D_i = X_i^T W_i^{-1} Y_i, \tag{6}$$

where D_i is a $m \times 1$ vector. Such transformations are familiar through traditional generalized least squares and in other Bayesian contexts. Expectations and covariances between the various quantities are as follows. Define

$$F_i = X_i^T W_i^{-1} X_i, \quad \text{and} \quad \bar{F} = \frac{1}{n} \sum_{i=1}^{n} F_i.$$

F_i and $\bar{F}$ are non-negative definite but not necessarily positive definite. We will stack the matrices $F_1, \ldots, F_n$ as the $nm \times m$ matrix $F^T = [F_1^T \; \ldots F_n^T]$. For expectations,

$$\text{E}(Y_i) = X_i \text{E}(\beta), \;\; \text{E}(D_i) = X_i^T W_i^{-1} X_i \text{E}(\beta) = F_i \text{E}(\beta).$$

For covariances,

$$\text{Var}(Y_i) = X_i \Gamma X_i^T + W_i, \;\; \text{Cov}(Y_i, Y_j) = X_i \Gamma X_j^T, \;\; \text{Var}(Y) = X \Gamma X^T + W,$$

for $i \neq j$, and where W is the direct sum $W = \oplus_{i=1}^n W_i$, i.e. the block diagonal matrix with diagonal elements $W_1, \ldots, W_n$. We have also

$$\text{Cov}(\beta, Y_i) = \Gamma X_i^T, \;\; \text{Cov}(\beta, Y) = \Gamma X^T.$$

Collect the constructed D_i quantities into the $nm \times 1$ vector $D^T = [D_1^T \; \dots \; D_n^T]$. We have

$$\mathrm{Var}(D_i) = F_i \Gamma F_i + F_i, \quad \mathrm{Cov}(D_i, D_j) = F_i \Gamma F_j, \mathrm{Var}(D) = F \Gamma F^T + \oplus_{i=1}^n F_i,$$

where $i \neq j$, and

$$\mathrm{Cov}(\beta, D_i) = \Gamma F_i, \quad \mathrm{Cov}(\beta, D) = \Gamma F^T.$$

Finally we have

$$\mathrm{Cov}(D_i, Y_i) = (F_i \Gamma + I_m) X_i^T = (\Gamma^{-1} + F_i) \Gamma X_i^T, \quad \mathrm{Cov}(D, Y) = F \Gamma X^T + \oplus_{i=1}^n X_i^T.$$

Some of the constructed variance matrices are not necessarily full rank. As such, the following results are useful.

Lemma 2.

$$\mathrm{Var}(A)\mathrm{Var}(A)^\dagger [A - \mathrm{E}(A)] = A - \mathrm{E}(A)$$

for any vector A of random quantities with coherent belief specifications and data consistent with them. Equivalently, $[A - \mathrm{E}(A)] \in \mathbf{range}\{\mathrm{Var}(A)\}$.

Lemma 3.

$$\mathrm{Var}(A)\mathrm{Var}(A)^\dagger \mathrm{Cov}(A, B) = \mathrm{Cov}(A, B)$$

for any vectors A, B of random quantities with coherent belief specifications. Equivalently, we have $\mathrm{Cov}(A, B) \in \mathbf{range}\{\mathrm{Var}(A)\}$.

Lemma 2 and Lemma 3 follow from Lemma 2.2.4 of Rao and Mitra (1971) and because coherent expectations and covariances must lie in the range of the corresponding variance matrix. Coherence of belief specifications, and subsequent data consistency are detailed in (Goldstein and Wooff 2007, Section 12.2).

Lemma 4. *Bartlett's identity (Bartlett 1951): for conformable matrices A, B, C, with A, B positive definite,*

$$(A + CB^{-1}C^T)^{-1} = A^{-1} - A^{-1}C(B + C^T A^{-1} C)^{-1} C^T A^{-1}.$$

See also the Sherman-Morrison-Woodbury formulae.

5 Bayes linear updating of regression parameters

We begin by noting formulae for the posterior expectations and variances for the parameter set β given the observables and the prior formulation. These are formally the appropriate adjusted expectations and variances within the BL paradigm, but also familiar within traditional linear Bayesian settings (Hartigan 1969; Mouchart and Simar 1980).

Lemma 5. *Given the observables, the BL revisions for the parameters β given single observations Y_i and all observations Y are as follows.*

$$\begin{aligned}
\mathrm{E}_{Y_i}(\beta) &= \mathrm{E}(\beta) + (\Gamma^{-1} + F_i)^{-1} X_i^T W_i^{-1} [Y_i - \mathrm{E}(Y_i)] && (7)\\
&= \mathrm{E}(\beta) + (\Gamma^{-1} + F_i)^{-1} [D_i - \mathrm{E}(D_i)] && (8)\\
\mathrm{E}_{Y}(\beta) &= \mathrm{E}(\beta) + (\Gamma^{-1} + n\bar{F})^{-1} X^T W^{-1} [Y - \mathrm{E}(Y)] && (9)\\
&= \mathrm{E}(\beta) + (\Gamma^{-1} + n\bar{F})^{-1} [D - \mathrm{E}(D)] && (10)\\
\mathrm{Var}_{Y_i}(\beta) &= (\Gamma^{-1} + F_i)^{-1} && (11)\\
\mathrm{Var}_{Y}(\beta) &= (\Gamma^{-1} + n\bar{F})^{-1}. && (12)
\end{aligned}$$

These follow by (1) and (2) and because we may write, using Lemma 4,

$$\mathrm{Var}(Y)^{-1} = W^{-1} - W^{-1} X (\Gamma^{-1} + n\bar{F})^{-1} X^T W^{-1}.$$

Notice in particular that the adjusted expectations (7) have a natural representation in terms of the transformed quantities $D_i = X_i^T W_i^{-1} Y_i$, which raises the question: what is the role of these quantities?

Theorem 6. *$\lfloor \beta \perp\!\!\!\perp Y_i \rfloor \,/\, D_i$ and the collection D_i is BL sufficient for Y_i for adjusting β.*

Proof. For any random vector A we must have, by Lemma 3,

$$\mathrm{Var}(D_i)\mathrm{Var}(D_i)^{\dagger}\mathrm{Cov}(D_i, A) = \mathrm{Cov}(D_i, A).$$

Thus,

$$\begin{aligned}
(F_i \Gamma F_i + F_i)(F_i \Gamma F_i + F_i)^{\dagger}\mathrm{Cov}(D_i, A) &= \mathrm{Cov}(D_i, A)\\
\Rightarrow (F_i + \Gamma^{-1})\Gamma F_i (F_i \Gamma F_i + F_i)^{\dagger}\mathrm{Cov}(D_i, A) &= \mathrm{Cov}(D_i, A)\\
\Rightarrow \Gamma F_i (F_i \Gamma F_i + F_i)^{\dagger}\mathrm{Cov}(D_i, A) &= (F_i + \Gamma^{-1})^{-1}\mathrm{Cov}(D_i, A). \qquad (13)
\end{aligned}$$

Now observe that

$$\begin{aligned}
\mathrm{Cov}(\beta, D_i)\mathrm{Var}(D_i)^{\dagger}\mathrm{Cov}(D_i, Y_i) &= \Gamma F_i (F_i \Gamma F_i + F_i)^{\dagger}\mathrm{Cov}(D_i, Y_i)\\
&= (F_i + \Gamma^{-1})^{-1}\mathrm{Cov}(D_i, Y_i) \text{ by (13)}\\
&= (F_i + \Gamma^{-1})^{-1}(I + F_i\Gamma)X_i^T\\
&= \Gamma X_i^T = \mathrm{Cov}(\beta, Y_i).
\end{aligned}$$

It follows by Lemma 1 that $\lfloor \beta \perp\!\!\!\perp Y_i \rfloor \,/\, D_i$, and this is necessary and sufficient to prove the result. □

The consequence is that we may update β either directly using Y_i or via the construct D_i.

Corollary 7. *$\lfloor D_i \perp\!\!\!\perp D_j \rfloor \,/\, \beta$ and $\lfloor D_i \perp\!\!\!\perp Y_j \rfloor \,/\, \beta$.*

Proof.

$$\mathrm{Cov}(D_i, \beta)\mathrm{Var}(\beta)^{-1}\mathrm{Cov}(\beta, D_j) = F_i^T \Gamma \Gamma^{-1} \Gamma F_j = F_i^T \Gamma F_j = \mathrm{Cov}(D_i, D_j),$$

so that $\lfloor D_i \perp\!\!\!\perp D_j \rfloor \,/\, \beta$ by Lemma 1. The second result follows similarly. □

Corollary 8. *The sample of transformed observables D is BL sufficient for the original observables Y for adjusting β, $\lfloor \beta \perp\!\!\!\perp Y \rfloor \,/\, D$.*

Proof. $\lfloor Y_i \perp\!\!\!\perp Y_j \rfloor \,/\, \beta$ for all $i \neq j$ and $\lfloor Y_i \perp\!\!\!\perp \beta \rfloor \,/\, D_i$ for all i by Theorem 6, and the result follows. □

5.1 Bayes linear sufficiency of the mean transformation

The sample mean vector $\bar{Y}$ of the sequence of observables $Y_1, Y_2, \ldots$ is not BL sufficient for Y for adjusting β. However, we now show that the transformed quantities D possess exploitable sufficiency properties. Begin by constructing the mean vector

$$\bar{D} = \frac{1}{n}\sum_{i=1}^{n} D_i = \bar{F}\beta + \frac{1}{n}\sum_{i=1}^{n} X_i^T W_i^{-1} \epsilon_i,$$

and note, following Section 4, the consequent specifications and covariance structures:

$$\mathrm{E}(\bar{D}) = \bar{F}\mathrm{E}(\beta), \;\; \mathrm{Cov}(\bar{D}, \beta) = \bar{F}\Gamma, \;\; \mathrm{Var}(\bar{D}) = (\bar{F}\Gamma\bar{F} + n^{-1}\bar{F}); \tag{14}$$

$$\mathrm{Cov}(\bar{D}, D_i) = \bar{F}\Gamma F_i + n^{-1}F_i, \;\; \mathrm{Cov}(\bar{D}, D) = \bar{F}\Gamma F^T + n^{-1}F^T. \tag{15}$$

Theorem 9. *The mean vector $\bar{D}$ is BL sufficient for D for adjusting β, and $\lfloor \beta \perp\!\!\!\perp D \rfloor / \bar{D}$.*

Proof. The proof is very similar to that of Theorem 6, with D_i and A there being replaced by $\bar{D}$ and D here. So, by Lemma 3,

$$\begin{aligned}
\mathrm{Var}(\bar{D})\mathrm{Var}(\bar{D})^{\dagger}\mathrm{Cov}(\bar{D}, D) &= \mathrm{Cov}(\bar{D}, D) \\
\Rightarrow (\bar{F}\Gamma\bar{F} + n^{-1}\bar{F})(\bar{F}\Gamma\bar{F} + n^{-1}\bar{F})^{\dagger}\mathrm{Cov}(\bar{D}, D) &= \mathrm{Cov}(\bar{D}, D) \\
\Rightarrow \Gamma\bar{F}(\bar{F}\Gamma\bar{F} + n^{-1}\bar{F})^{\dagger}\mathrm{Cov}(\bar{D}, D) &= (\bar{F} + n^{-1}\Gamma^{-1})^{-1}\mathrm{Cov}(\bar{D}, D).
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathrm{Cov}(\beta, \bar{D})\mathrm{Var}(\bar{D})^{\dagger}\mathrm{Cov}(\bar{D}, D) &= \Gamma\bar{F}[(\bar{F}\Gamma\bar{F} + n^{-1}\bar{F})^{\dagger}]\mathrm{Cov}(\bar{D}, D) \\
&= (\bar{F} + n^{-1}\Gamma^{-1})^{-1}(\bar{F}\Gamma F^T + n^{-1}F^T) \\
&= \Gamma F^T = \mathrm{Cov}(\beta, D).
\end{aligned}$$

The result follows by Lemma 1. □

Corollary 10. $\lfloor \beta \perp\!\!\!\perp Y \rfloor \,/\, \bar{D}$.

Proof. Follows similarly. □

The consequence is that we may update β directly using the sequence $Y_1, Y_2, \ldots, Y_n$ or indirectly via the average $\bar{D}$ of the transformed values of Y_i.

6 Prediction of future observables

We have shown that $\lfloor \beta \perp\!\!\!\perp Y \rfloor \,/\, \bar{D}$, so that the transformed quantities are BL sufficient for the parameter collection. We now extend this notion to prediction for a further collection of observables, $Y_{n+1}, Y_{n+2}, \ldots, Y_{n+r}$. The following theorem is central.

Theorem 11. *Partition the observables into two disjoint subsets, Y^* and Y^{**}. Construct the corresponding disjoint subsets of transformed quantities, D^* and D^{**}, and construct the mean vector $\bar{D}^*$ for the first set. Then $\bar{D}^*$ is BL sufficient for Y^* for adjusting Y^{**}, and $\lfloor Y^{**} \perp\!\!\!\perp Y^* \rfloor \,/\, \bar{D}^*$.*

Proof. For any $r > n$ and any $1 \leq i \leq n$, we may use (13) to show that

$$\begin{aligned}\mathrm{Cov}(D_r, \bar{D})\mathrm{Var}(\bar{D})^{\dagger}\mathrm{Cov}(\bar{D}, D_i) &= F_r \Gamma \bar{F}(\bar{F}\Gamma\bar{F} + n^{-1}\bar{F})^{\dagger}\mathrm{Cov}(\bar{D}, D_i)\\ &= F_r \Gamma F_i = \mathrm{Cov}(D_r, D_i).\end{aligned}$$

It follows that $\lfloor D_r \perp\!\!\!\perp D_i \rfloor \,/\, \bar{D}$, and thence to $\lfloor D^{**} \perp\!\!\!\perp D^* \rfloor \,/\, \bar{D}^*$. The main result follows similarly. □

Corollary 12. *The adjusted expectation for β given $\bar{D}^*$, $\mathrm{E}_{\bar{D}^*}(\beta)$ is BL sufficient for Y^* for adjusting Y^{**}, and $\lfloor Y^{**} \perp\!\!\!\perp Y^* \rfloor \,/\, \mathrm{E}_{\bar{D}^*}(\beta)$.*

The main implication is that if we wish to revise our beliefs about a second set of observables by a disjoint set of observables, we may do so via updating the intermediary parameter set β by the mean of the first set of transformed observables.

7 Diagnostics, residuals, and sequential analysis

The parameter set β is m-dimensional, whereas the space of observables Y is nk-dimensional. Linear fitting of β onto the linear space $\langle Y \rangle$ implies (at most) an m-dimensional subspace of $\langle Y \rangle$ which is informative for β, and this subspace can now be seen to be identified with $\langle \bar{D} \rangle$. For general BL updating, this subspace is termed the *heart of the transform*, denoted $\mathbb{H}(Y/\beta)$=$\mathbb{H}(\bar{D}/\beta)$, and it may be constructed via the eigenstructure of the reverse transform $\mathbb{T}_{Y:\beta}$ (Goldstein and Wooff 2007). The remaining $(nk - m)$-dimensional subspace is denoted $\mathbb{H}^{\perp}(Y/\beta)$, here being the orthogonal complement of $\mathbb{H}(\bar{D}/\beta)$ in $\langle Y \rangle$. This is an *ancillary* space: it can tell us nothing about the parameter collection β, but exploration of it might allow us to diagnose problems with our prior formulation. In order to do this, we may construct alternative sets of residuals. In terms of the original observables, the appropriate residuals are the quantities $Y_i - \mathrm{E}_{\bar{D}}(Y_i)$, which span the ancillary space $\mathbb{H}^{\perp}(Y/\beta)$. It may also be valuable to inspect the residuals of the transformed observables from their mean: $D_i - \mathrm{E}_{\bar{D}}(D_i)$ in order to detect anomalies in the transformed space of pseudo-exchangeables.

For problems with a natural time-ordering, evaluation of sequential diagnostics is appropriate as follows. We update expectations for the parameter vector β by observations as they arrive. By Corollary 12, the adjusted expectation for β given $\bar{D}^*$ is BL sufficient for the early observations Y^* for adjusting later observations Y^{**}. Consequently, the time-order residuals of interest are the vectors $Y_r - \mathrm{E}_{\mathrm{E}_{\bar{D}^*}(\beta)}(Y_r)$, for $Y_r \in Y^{**}$. We discuss in Goldstein and Wooff (2007) how we may use residuals for variance learning. It is then possible to employ updated variance components within two-stage BL analysis.

7.1 Example: multivariate regression with correlated errors

Figure 4 shows a BL influence diagram (Goldstein and Wooff 1995) produced by the software package [B/D] (Wooff and Goldstein 2000). For interpreting such diagrams, see Goldstein and Wooff (2007). Shown are two columns of residuals computed via sufficiency. The left-hand column of residuals shows diagnostics for the simple two-dimensional vector residuals $R_i = Y_i - \mathrm{E}_{\bar{D}}(Y_i)$, i.e. the residuals conditioning on the full sufficient information $\bar{D}$. The right-hand column of residuals shows the sequential revision of the observable quantities as fresh information arises. That is, we exploit sufficiency to compute $L_i = \mathrm{E}_{\bar{D}_{1,\ldots,i}}(\beta)$, the posterior vector of expectations for the regression coefficients, given all observables up to and including observation Y_i. The sequential residuals of interest are the vectors $\tilde{R}_i = Y_i - \mathrm{E}_{L_i}(Y_i)$. The first residual pair is the same, $R_1 = \tilde{R}_1$.

There are many diagnostic features we might want to assess, including features in the ancillary space. Typically our residual analyses relate back to the graphical templates constructed over the duplicated structures of these multivariate multiple regression problems, however there is only scope here for a brief illustrative graphical assessment of simple residuals and temporal behaviour. Among the features we see on the plot, $\bar{D}$ is 4-dimensional and sufficient for the 12-dimensional observation space $\langle Y \rangle$. Once observed it is used to update the parameter set β, which has 76.4% of its uncertainty explained. The preponderance of dark shading implies that the data produce quite large revisions of expectation relative to variation explained. The simple residuals R_i show different patterns of substantial light and dark shading. Light shading implies unexpectedly small revisions in expectation relative to variation explained. We tend to see unexpectedly large revisions in early and late time, and unexpectedly small revisions in middle time. These residuals should exhibit random diagnostic patterns, so a conclusion might be that temporal assumptions about the error quantities are inappropriate.

The sequential residuals $\tilde{R}_i$ are used to assess temporal features in the revision process, conditioning partially on the sufficient information as it arrives. Take for example the node summarising $\tilde{R}_4$ as information arrives. The node is divided into four sectors,

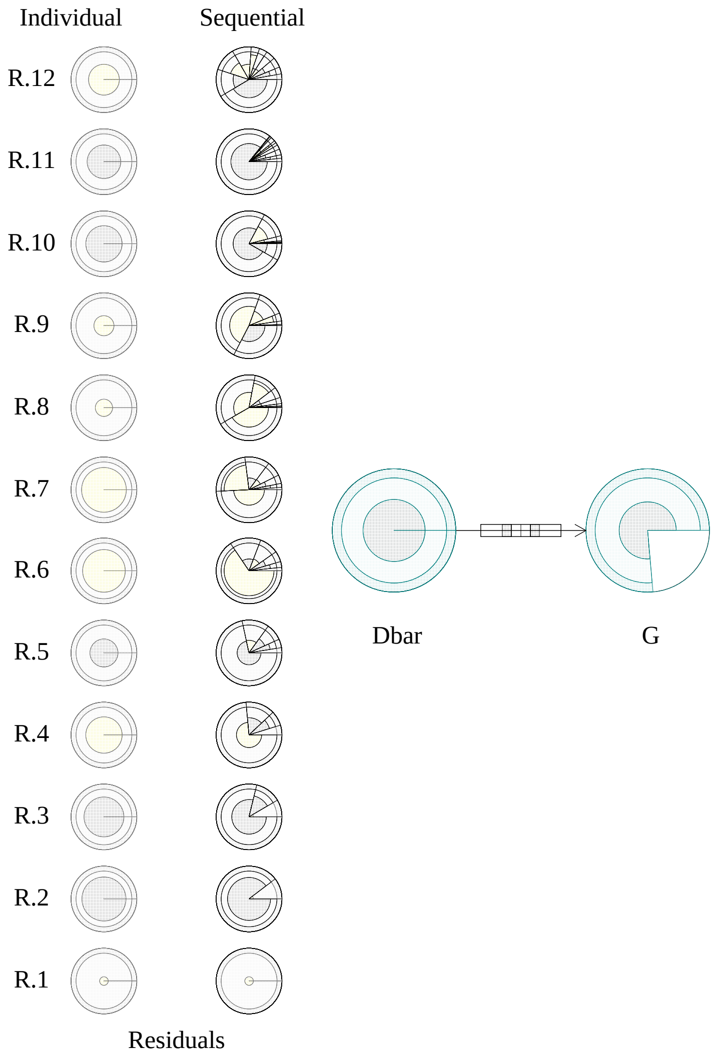

Figure 4: Alternative diagnostic analyses produced by the BL computation package [B/D]. G corresponds to β; *Dbar* to $\bar{D}$, and $R.i$ to residual diagnostics.

starting at 0^0 and moving anti-clockwise. These provide summary diagnostics, in sequence, for the revision of Y_4 given the sufficient evidence $\bar{D}_1$ at time T_1; diagnostics for the partial revision of Y_4 by new evidence $\bar{D}_2$ given $\bar{D}_1$, diagnostics for the partial revision of Y_4 by new evidence $\bar{D}_3$ given $\bar{D}_{1,2}$, and finally diagnostics for the observation Y_4 given new evidence $\bar{D}_4$ given $\bar{D}_{1,2,3}$. The arc lengths of sectors correspond to percentages of variation explained at each stage. For Y_4 we observe that the information available at T_1 is barely relevant. By T_3 around 25% of variation has been explained, but with larger revisions than anticipated. At T_4 any remaining variation is explained, but the light shading implies that the observation was much closer to its forecast then expected. This is mostly the case for other residuals $\tilde{R}_i$ in middle time. Comparing the two sets of residuals we see that the partial diagnostics tend to agree with those for the simple residuals, suggesting that any aberrant behaviour in the revision process for Y_i can be attributed to aspects of its model and observation and not to other observations.

8 Parameters dependent on design point

We consider now extending to multiple multivariate regressions which do not share a common parameter space, i.e. we replace β by β_i and so have n separate multiple regressions, each of the form $Y_i = X_i\beta_i + \epsilon_i$. Generally, there are no sufficiencies to exploit. However, there are situations where the model parameters are specific to a design point, but related via a simple correlation structure. Such models are employed, for example, in history matching of oil reservoirs. For vectors $Y_i, \beta_i, \epsilon_i, \delta_i$ and design matrices X_i, $i = 1 \ldots n$, consider the set of models

$$Y_i = X_i\beta_i + \epsilon_i, \quad \text{where} \quad \beta_i = \tilde{\beta} + \delta_i,$$

with $\mathrm{E}(\tilde{\beta}) = \mathrm{E}(\beta)$, $\mathrm{Var}(\tilde{\beta}) = \Gamma$, and where $\delta_1, \ldots, \delta_n$ is an uncorrelated sequence of parameter perturbations, defined to be uncorrelated with β, and with $\mathrm{Var}(\delta_i) = \tilde{V}_i$. This special case also includes the situation in which the β_i vectors are themselves exchangeable, in which case $\tilde{V}_i$ is the same for all i. We may then express

$$Y_i = X_i\tilde{\beta} + \tilde{\epsilon}_i, \tag{16}$$

where $\tilde{\epsilon}_i = \epsilon_i + X_i\delta_i$ and $\mathrm{Var}(\tilde{\epsilon}_i)$=$W_i + X_i\tilde{V}_iX_i^T = \tilde{W}_i$. This is functionally identical to the general model considered herein, and so can be treated in the same way, with the sufficiencies depending on a transform of the error variance structure as well as the observables.

9 Implied exchangeability and sample size design

We now show how to use the sufficiencies we have identified to drive sample-size design. Suppose we return to our example of Section 3. Suppose we ask two questions. First, how many observations must we obtain in order to achieve a specified precision in a particular linear combination of the parameters β? Secondly, where should we design subsequent observations? The first of these questions can be answered at least approximately as follows.

Theorem 13. *The sequence of observables $Y_1, Y_2, \ldots, Y_n$ and its implications for learning about β is consistent with the existence of an infinite second-order exchangeable sequence $C_1, C_2, \ldots$ and the use of this sequence for updating beliefs about its underlying mean $\mathcal{A}$.*

Proof. Construct an infinite second-order exchangeable series of vectors $C_1, C_2, \ldots$, with $C_i = \mathcal{A} + \mathcal{S}_i$, such that prior beliefs concerning the sequence $C_1, C_2, \ldots$ are as follows. For all i and $j \neq i$,

$$\mathrm{E}(\mathcal{A}) = \mathrm{E}(\beta),\ \mathrm{E}(\mathcal{S}_i) = 0,\ \mathrm{Var}(\mathcal{A}) = \Gamma,\ \mathrm{Var}(\mathcal{S}_i) = \bar{F},\ \mathrm{Cov}(\mathcal{S}_i, \mathcal{A}) = 0,\ \mathrm{Cov}(\mathcal{S}_i, \mathcal{S}_j) = 0.$$

It is readily established that the revision of belief over β using data $Y_1, \ldots, Y_n$ matches the revision of belief over $\mathcal{A}$ using "data" $C_1, C_2, \ldots$, because

$$\mathrm{E}_Y(\beta) = \mathrm{E}_C(\mathcal{A}),\quad \mathrm{Var}_Y(\beta) = \mathrm{Var}_C(\mathcal{A}),\quad \text{and}\quad \mathbb{T}_{\beta:Y} = \mathbb{T}_{\mathcal{A}:C}.$$

□

We shall call $C_1, C_2, \ldots$ the implied infinite exchangeable sequence associated with the observed non-exchangeable sequence $Y_1, Y_2, \ldots, Y_n$. We now exploit the implied sequence for sample-size design.

9.1 Approximate sample-size design

The study of sample-size implications for infinite exchangeable sequences is based on the canonical form, considered in Goldstein and Wooff (1997, 1998) and, for finite exchangeable sequences, Shaw and Goldstein (2012). Thus, when we may construct an implied exchangeable sequence as above, the resolution transform matrix $\mathbb{T}_{\mathcal{A}:C}$ may be used to explore sample size considerations, on the assumption that the mean weighted precision matrix $\bar{F}$ is "typical" of the weighted precision matrices

$$F_{n+1} = X_{n+1}^T W_{n+1}^{-1} X_{n+1},\ \ F_{n+2} = X_{n+2}^T W_{n+2}^{-1} X_{n+2} \ldots.$$

for further observations. As an example, Figure 5 shows the implications of increasing the number of observations beyond $n = 12$, assuming that the weighted precision matrices for subsequent design points are typical of those encountered so far. Variance resolutions for linear combinations $h^T\beta$ of the parameter set are bounded by the maximal and minimal eigenvalues, $\lambda_{max}, \lambda_{min}$, of $\mathbb{T}_{\mathcal{A}:C}$. The plot shows that further design points would contribute only marginally to resolving the present residual uncertainty in the parameter collection. It can be shown that the direction of maximal variance reduction is approximately $\beta_3 - \beta_1$, and the direction of minimal reduction is approximately $\beta_3 + 2\beta_1$. For details and guidance on sample-size features for exchangeable and co-exchangeable sequences, see Goldstein and Wooff (2007).

9.2 Deduction of appopriate linear transformation

We may apply the notion in reverse. Suppose that $Q = Q_1, \ldots, Q_n$ are any set of observables linearly related to a parameter collection β, with $\mathrm{Var}(\beta) = \Gamma$, where Γ is positive definite, and suppose that it turns out that the posterior variance $\mathrm{Var}_Q(\beta)$ is of the form

$$\mathrm{Var}_Q(\beta) = (\Gamma^{-1} + nG)^{-1}$$

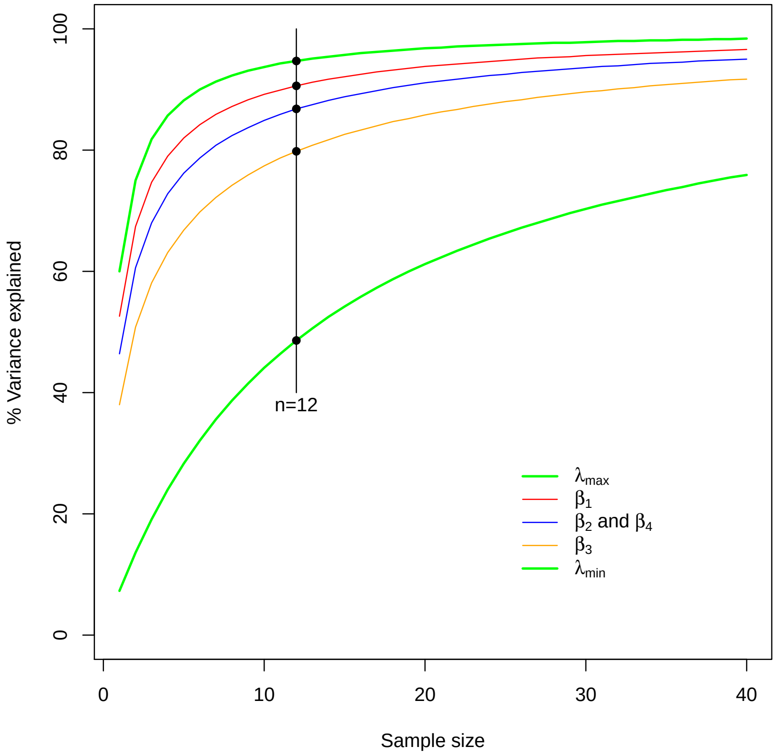

Figure 5: The implication of taking different sample sizes for explaining variance in the regression parameters. Actual sample size is marked at $n = 12$.

for some non-negative definite matrix G which does not depend on n. Then there exists an implied infinite second-order exchangeable sequence $C_1, C_2, \ldots$ with properties as above. There may then exist linear transformations $Q_i' = P_iQ_i$ for some $P_1, \ldots, P_n$ such that $\lfloor Q_i \perp\!\!\!\perp \beta \rfloor / Q_i'$ and such that the mean $\bar{Q}'$ is BL sufficient for the original observables for adjusting β. This suggests that if our starting position is a variance matrix with such standard form, we should be able to deduce which linear transformations P_i of the original observables to take in order to produce BL sufficient transformations.

9.3 Design of observations

To discuss the second question we posed, the idea that all multivariate multiple regressions of the forms considered may be associated with an implied infinite exchangeable sequence has interesting implications for sequential design of further observations for non-excheangeable multivariate multiple regressions. This problem is already difficult in the exchangeable setting: see, for example, Pilz (1991). That is, given n observations, choose the next design matrix X_{n+1} to satisfy criteria such as maximising the portion of residual variation in the parameter set explained by the additional observation. This will be a focus of future work.

10 Discussion

The identification of sufficiency vastly simplifies computations and interpretations for BL and approximate Bayesian inference for high-dimensional models. Further, the associated residual structures form natural foci for diagnostic assessment, and the association of the regression problem with an implied infinite exchangeable sequence allows tractable sample-size design even for very large scale problems.

The separation of information Y_i into a transformed sufficient portion D_i, together with other identified sufficiencies and different kinds of residual structure, has an implication for the graphical models and plates which motivated this paper. In particular, in future work we wish to propose different kinds of operations on such graphs which will make plain the relationships between these components. This will clarify whether there are further implications for the junction tree and for local computation (Goldstein and Wilkinson 2000) over the graph. Further, these operations should allow us to focus more on alternative ways of representing the ancillary space, namely the components which are not useful in making revisions over the parameter set or in making predictions, but which are useful in making diagnostic assessments of the prior formulation.

Acknowledgements

I am grateful to two referees for their comments and their different perspectives which I feel have led to a better paper.

References

Bartlett, M. S. (1951). “An inverse matrix adjustment arising in discriminant analysis.” *Ann. Math. Statist.*, 22: 107. 85

Blum, M. G. B., Nunes, M. A., Prangle, D., and Sisson, S. A. (2013). “A Comparative Review of Dimension Reduction Methods in Approximate Bayesian Computation.” *Statistical Science*, 28(2): 189–208. 80

Box, G. E. P. and Tiao, G. C. (1973). *Bayesian inference in statistical analysis*. Addison-Wesley. 80

Cumming, J. A. and Goldstein, M. (2009). “Small Sample Bayesian Designs for Complex High-Dimensional Models Based on Information Gained Using Fast Approximations.” *Technometrics*, 51(4): 377–388. 78

de Finetti, B. (1937). “Foresight:its logical laws, its subjective sources.” *Annales de l’Institut Henri Poincaré*, 7. 78

— (1974). *Theory of probability, vol. 1*. New York: Wiley. 78

Goldstein, M. (1981). "Revising previsions: a geometric interpretation." *J. R. Statist. Soc.*, B:43: 105–130. 78

— (1986). "Separating beliefs." In Goel, P. and Zellner, A. (eds.), *Bayesian Inference and Decision Techniques*. Amsterdam: North Holland. 80

— (1990). "Influence and belief adjustment." In Smith, J. and Oliver, R. (eds.), *Influence Diagrams, Belief Nets and Decision Analysis*. Chichester: Wiley. 80

Goldstein, M. and Rougier, J. C. (2006). "Bayes Linear Calibrated Prediction for Complex Systems." *J. Am. Statist, Assoc.*, 101(475): 1132–1143. 78

— (2009). "Reified Bayesian Modelling and Inference for Physical Systems, with discussion and rejoinder." *J. Statist. Plan. Inf.*, 139(3): 1221–1239. 78

Goldstein, M. and Wilkinson, D. J. (2000). "Bayes linear analysis for graphical models: the geometric approach to local computation and interpretive graphics." *Statistics and Computing*, 10(4): 311–324. 80, 94

Goldstein, M. and Wooff, D. A. (1995). "Bayes linear computation: concepts, implementation and programming environment." *Statistics and Computing*, 5: 327–341. 89

— (1997). "Choosing sample sizes in balanced experimental designs: a Bayes linear approach." *The Statistician*, 46(2): 167–183. 92

— (1998). "Adjusting exchangeable beliefs." *Biometrika*, 85(1): 39–54. 80, 92

— (2007). *Bayes linear statistics: Theory and methods.*. Chichester: Wiley. 78, 80, 81, 82, 85, 88, 89, 92

Gosling, J. P., Hart, A., Owen, H., Davies, M., Li, J., and MacKay, C. (2013). "A Bayes Linear Approach to Weight-of-Evidence Risk Assessment for Skin Allergy." *Bayesian Analysis*, 8(1): 169–186. 78

Hartigan, J. A. (1969). "Linear Bayes methods." *J. Roy. Statist. Soc., B*, 31: 446–454. 85

Mouchart, M. and Simar, L. (1980). "Least squares approximation in Bayesian analysis (with Discussion)." In Bernardo, J.-M. et al. (eds.), *Bayesian Statistics*, 207–222. University Press, Valencia, Spain. 85

Pilz, J. (1991). *Bayesian estimation and experimental design in linear regression models*. Chichester: Wiley. 93

Rao, C. R. and Mitra, S. K. (1971). *Generalized inverse of matrices and its applications*. New York: Wiley. 85

Rougier, J. C. and Kern, M. (2010). "Predicting snow velocity in large chute flows under different environmental conditions." *J. Roy. Statist. Soc. C: Applied Statistics*, 59(5): 737–760. 78

Shaw, S. C. and Goldstein, M. (2012). "Finite population corrections for multivariate Bayes sampling." *Journal of Statistical Planning and Inference*, 42(10): 2844–2861. 92

Vernon, I. R., Goldstein, M., and Bowers, R. G. (2010). "Galaxy Formation: a Bayesian Uncertainty Analysis (with discussion)." *Bayesian Analysis*, 5(4): 619–708. 78

Wilkinson, D. J. (1998). "An object-oriented approach to local computation in Bayes linear belief networks." In Green, P. J. and Payne, R. W. (eds.), *Proceedings in computational statistics*, 491–496. Heidelberg: Physica Verlag. 80

Williamson, D. E., Goldstein, M., and Blaker, A. (2012). "Fast Linked Analyses for Scenario based Hierarchies." *J. Roy. Statist. Soc. C: Applied Statistics*, 61(5): 665–691. 78

Wooff, D. A. and Goldstein, M. (2000). "The Bayes linear programming language [B/D]." *Journal of Statistical Software*, 5(2). Http://www.stat.ucla.edu/journals/jss/v05/i02. 89