Introduction to Bayes Linear Statistics

Jonathan Cumming, Ian Vernon

3rd September, 2010
Introduction
Bayesian Inference

- **Bayesian inference** is a method of statistical inference in which we use data to update beliefs about uncertain quantities of interest.
Bayesian Inference

- **Bayesian inference** is a method of statistical inference in which we use data to update beliefs about uncertain quantities of interest.

- In this methodology, our uncertainty about any quantities of interest is quantified by probability distributions.
Bayesian Inference

- **Bayesian inference** is a method of statistical inference in which we use data to update beliefs about uncertain quantities of interest.

- In this methodology, our uncertainty about any quantities of interest is quantified by **probability distributions**.

- Our updated beliefs about the quantity of interest, $\theta$, given the data, $D$, are then obtained via application of **Bayes Theorem**:

$$ p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} $$
Bayesian Inference

- \( p(\theta) \) represents our beliefs about \( \theta \) before the data \( D \) become available - the prior for \( \theta \).
Bayesian Inference

- \( p(\theta) \) represents our beliefs about \( \theta \) before the data \( D \) become available - the prior for \( \theta \).
- \( p(D|\theta) \) is the probability of the data given the uncertain quantity \( \theta \).
Bayesian Inference

- $p(\theta)$ represents our beliefs about $\theta$ before the data $D$ become available - the prior for $\theta$.
- $p(D|\theta)$ is the probability of the data given the uncertain quantity $\theta$.
- $p(\theta|D)$ represents our beliefs about $\theta$ after the data $D$ have been observed - the posterior for $\theta$. 
Bayesian Inference

- $p(\theta)$ represents our beliefs about $\theta$ before the data $D$ become available - the prior for $\theta$.
- $p(D|\theta)$ is the probability of the data given the uncertain quantity $\theta$.
- $p(\theta|D)$ represents our beliefs about $\theta$ after the data $D$ have been observed - the posterior for $\theta$.
- These first three elements are the heart of Bayesian inference, which can be remembered as

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior}$$
Some interesting questions

- We need to specify a full joint probability distribution for all uncertain quantities
  - Specifying a probability distribution is equivalent to specifying an infinite set of moments, do we really believe all of those implicit specifications?
Some interesting questions

- We need to specify a full joint probability distribution for all uncertain quantities
  - Specifying a probability distribution is equivalent to specifying an infinite set of moments, do we really believe all of those implicit specifications?
- When the prior and likelihood have specific forms, then this posterior can be found analytically (conjugacy)
  - Are those distributions really REALLY conjugate, or are they just convenient?
Some interesting questions

- We need to specify a full joint probability distribution for all uncertain quantities
  - Specifying a probability distribution is equivalent to specifying an infinite set of moments, do we really believe all of those implicit specifications?

- When the prior and likelihood have specific forms, then this posterior can be found analytically (conjugacy)
  - Are those distributions really REALLY conjugate, or are they just convenient?

- In all other cases, we must rely on intensive computational methods to arrive a distribution for $p(\theta|D)$
  - If we don’t completely believe our prior specification, what faith should we have in this posterior?
Difficulties with full Bayes

- Working within a fully Bayesian framework, we can encounter certain difficulties when considering multivariate analyses.
Difficulties with full Bayes

- Working within a fully Bayesian framework, we can encounter certain difficulties when considering multivariate analyses.
- Even in small problems, it can be extremely difficult and/or time-consuming to distil all the prior knowledge into a meaningful joint prior probability specification.
Difficulties with full Bayes

- Working within a fully Bayesian framework, we can encounter certain difficulties when considering multivariate analyses.
- Even in small problems, it can be extremely difficult and/or time-consuming to distil all the prior knowledge into a meaningful joint prior probability specification.
- Even with a specification, the computations for learning from data become both difficult and computer intensive.
Difficulties with full Bayes

- Working within a fully Bayesian framework, we can encounter certain difficulties when considering multivariate analyses.
- Even in small problems, it can be extremely difficult and/or time-consuming to distil all the prior knowledge into a meaningful joint prior probability specification;  
- Even with a specification, the computations for learning from data become both difficult and computer intensive; 
- In higher-dimensions the likelihood surface can be very complicated, making full Bayes calculations potentially highly non-robust.
Difficulties with full Bayes

- Working within a fully Bayesian framework, we can encounter certain difficulties when considering multivariate analyses.
- Even in small problems, it can be extremely difficult and/or time-consuming to distil all the prior knowledge into a meaningful joint prior probability specification;
- Even with a specification, the computations for learning from data become both difficult and computer intensive;
- In higher-dimensions the likelihood surface can be very complicated, making full Bayes calculations potentially highly non-robust.
- Therefore if we are unable to make and analyse full prior probability specifications, we require methods based around simpler belief specifications.
Expectation as Primitive

- Rather than work with probability as the fundamental quantity of uncertainty, we could use expectation.

- de Finetti spent most of his life studying subjective conceptions of probability.

- He proposed the use of expectation as the primitive entity on which to base any analysis, as opposed to probability.
Expectation as Primitive

- Rather than work with probability as the fundamental quantity of uncertainty, we could use expectation.
- de Finetti spent most of his life studying subjective conceptions of probability.
- He proposed the use of expectation as the primitive entity on which to base any analysis, as opposed to probability.
- Probabilities (where relevant) enter as derived quantities: they are the expectations of indicator functions.
Expectation as Primitive

- Rather than work with probability as the fundamental quantity of uncertainty, we could use expectation.
- de Finetti spent most of his life studying subjective conceptions of probability.
- He proposed the use of expectation as the primitive entity on which to base any analysis, as opposed to probability.
- Probabilities (where relevant) enter as derived quantities: they are the expectations of indicator functions.
- Note this asymmetry: if probability is treated as the primitive quantity then one has to specify (in the continuous case) an infinite set of probabilities in order to derive a single expectation.
Expectation as Primitive

- Rather than work with probability as the fundamental quantity of uncertainty, we could use expectation.
- de Finetti spent most of his life studying subjective conceptions of probability.
- He proposed the use of expectation as the primitive entity on which to base any analysis, as opposed to probability.
- Probabilities (where relevant) enter as derived quantities: they are the expectations of indicator functions.
- Note this asymmetry: if probability is treated as the primitive quantity then one has to specify (in the continuous case) an infinite set of probabilities in order to derive a single expectation.
Introduction to Bayes Linear Statistics

Jonathan Cumming, Ian Vernon
In the Bayes linear approach, we follow de Finetti and take expectation as primitive.
Working with partial belief specifications

- In the Bayes linear approach, we follow de Finetti and take expectation as primitive.

- We construct partial belief specifications using only means, variances and covariances for all uncertain quantities.
Working with partial belief specifications

- In the Bayes linear approach, we follow de Finetti and take expectation as primitive.
- We construct partial belief specifications using only means, variances and covariances for all uncertain quantities.
- We may view the Bayes linear approach as
  - Offering a simple approximation to a full Bayes analysis
  - Complementary to the full Bayes approach, offering new interpretative and diagnostic tools
  - A generalisation of the full Bayes approach where we lift the restriction of requiring a full probabilistic prior before we may learn anything from data
Features of the Bayes linear approach

- Subjective and Bayesian
Features of the Bayes linear approach

- Subjective and Bayesian
- Belief specifications *genuinely* correspond to our beliefs
Features of the Bayes linear approach

- Subjective and Bayesian
- Belief specifications *genuinely* correspond to our beliefs
- Expectation as *primitive*
Features of the Bayes linear approach

- Subjective and Bayesian
- Belief specifications **genuinely** correspond to our beliefs
- Expectation as **primitive**
- Adjust beliefs by **linear fitting** rather than conditioning
Features of the Bayes linear approach

- Subjective and Bayesian
- Belief specifications **genuinely** correspond to our beliefs
- Expectation as **primitive**
- Adjust beliefs by **linear fitting** rather than conditioning
- Computationally **straightforward** allowing the analysis of larger and more complex problems
Features of the Bayes linear approach

- Subjective and Bayesian
- Belief specifications *genuinely* correspond to our beliefs
- Expectation as *primitive*
- Adjust beliefs by *linear fitting* rather than conditioning
- Computationally *straightforward* allowing the analysis of larger and more complex problems
- **Diagnostic tools** are an important part of the approach
  - How prior beliefs affect conclusions
  - How beliefs change by the adjustment
  - How beliefs about observables compare to the observations themselves
### Features of the Bayes linear approach

- **Subjective and Bayesian**
- **Belief specifications** genuinely correspond to our beliefs
- **Expectation** as primitive
- Adjust beliefs by **linear fitting** rather than conditioning
- Computationally **straightforward** allowing the analysis of larger and more complex problems
- **Diagnostic tools** are an important part of the approach
  - How prior beliefs affect conclusions
  - How beliefs change by the adjustment
  - How beliefs about observables compare to the observations themselves
- **Important special cases** - multivariate Gaussian
Belief Specification

- The Bayes linear approach is **subjectivist**, and so in any analysis we need to specify our beliefs over all random quantities of interest.
Belief Specification

- The Bayes linear approach is subjectivist, and so in any analysis we need to specify our beliefs over all random quantities of interest.

- However, as we consider expectation as primitive we make our belief specifications in terms of the low-order moments of the random quantities of interest.
Belief Specification

- The Bayes linear approach is subjectivist, and so in any analysis we need to specify our beliefs over all random quantities of interest.

- However, as we consider expectation as primitive we make our belief specifications in terms of the low-order moments of the random quantities of interest.

- (If we have beliefs about higher orders we can include these in the analysis too)
Belief Specification

- The Bayes linear approach is subjectivist, and so in any analysis we need to specify our beliefs over all random quantities of interest.
- However, as we consider expectation as primitive we make our belief specifications in terms of the low-order moments of the random quantities of interest.
- (If we have beliefs about higher orders we can include these in the analysis too)
- For example, say we are interested in predicting $B = (B_1, B_2)^T$ from knowledge of $D = (D_1, D_2)^T$ which we will measure soon, then all we need to specify are $E(B)$, $E(D)$, $\text{Var}(B)$, $\text{Var}(D)$ and $\text{Cov}(B, D)$.
The example: Numbers

- Suppose we have a four quantities of interest, 
  \[ F = (B_1, B_2, D_1, D_2)^T \]
The example: Numbers

- Suppose we have a four quantities of interest, \( F = (B_1, B_2, D_1, D_2)^T \)
- We observe values of \( D = (D_1, D_2)^T \), and want to analyse the effects on our beliefs about \( B \)
The example: Numbers

- Suppose we have a four quantities of interest, $F = (B_1, B_2, D_1, D_2)^T$.
- We observe values of $D = (D_1, D_2)^T$, and want to analyse the effects on our beliefs about $B$.
- We have a very simple prior specification:

  $E(F)_i = 0, \quad \text{Var}(F)_{ii} = 100,$

and we have a correlation structure as follows:

\[
\begin{array}{cccc}
  B_1 & B_2 & D_1 & D_2 \\
  B_1 & 1.00 & 0.56 & 0.52 & 0.61 \\
  B_2 & 0.56 & 1.00 & 0.32 & 0.98 \\
  D_1 & 0.52 & 0.32 & 1.00 & 0.28 \\
  D_2 & 0.61 & 0.98 & 0.28 & 1.00 \\
\end{array}
\]
Stages of belief analysis

A typical Bayes linear analysis of beliefs proceeds in the following stages:

1. **Specification** of prior beliefs
2. Interpret the expected adjustments *a priori*
3. Given observations, perform and interpret the *adjustments*
4. Make *diagnostic comparisons* between actual and expected beliefs
Bayes Linear Inference
Belief Adjustment

- We are interested in how our beliefs about $B$ change in the light of information given by $D$. 
Belief Adjustment

- We are interested in how our beliefs about $B$ change in the light of information given by $D$.
- We look among the collection of linear estimates, i.e. those of form $c_0 + c_1 D_1 + c_2 D_2$, and choose constants $c_0, c_1, c_2$ to minimise the prior expected squared error loss in estimating each of $B_1$ and $B_2$:

$$\mathbb{E}([B_1 - c_0 - c_1 D_1 - c_2 D_2]^2).$$
Belief Adjustment

- We are interested in how our beliefs about $B$ change in the light of information given by $D$.
- We look among the collection of linear estimates, i.e. those of form $c_0 + c_1 D_1 + c_2 D_2$, and choose constants $c_0, c_1, c_2$ to minimise the prior expected squared error loss in estimating each of $B_1$ and $B_2$:

$$E([B_1 - c_0 - c_1 D_1 - c_2 D_2]^2).$$

- The choices of constants may be easily computed, and the estimators $E_D(B) = (E_D(B_1), E_D(B_2))^T$ turn out to be given by:

$$E_D(B) = E(B) + \text{Cov}(B, D) \text{Var}(D)\dagger(D - E(D)).$$

which we refer to as the adjusted expectation for collection $B$ given collection $D$. 
Adjusted expectation

- The adjusted expectation for collection $B$ given collection $D$ is

$$E_D(B) = E(B) + \text{Cov}(B, D) \text{Var}(D)^\dagger (D - E(D)).$$
Adjusted expectation

- The **adjusted expectation** for collection $B$ given collection $D$ is

$$E_D(B) = E(B) + \text{Cov}(B, D) \text{Var}(D)^\dagger(D - E(D)).$$

- The **adjusted version** of the $B$ given $D$ is the ‘residual’ vector

$$A_B(D) = B - E_D(B).$$
Adjusted expectation

- The adjusted expectation for collection $B$ given collection $D$ is

$$E_D(B) = E(B) + \text{Cov}(B, D) \text{Var}(D)^\dagger (D - E(D)).$$

- The adjusted version of the $B$ given $D$ is the ‘residual’ vector

$$A_B(D) = B - E_D(B).$$

- We can partition the vector $B$ as the sum of two uncorrelated vectors:

$$B = E_D(B) + A_B(D).$$
Adjusted variance

- We partition the variance matrix of $B$ into two variance components:

$$\text{Var}(B) = \text{Var}(E_D(B)) + \text{Var}(A_B(D))$$

$$= \text{RVar}_D(B) + \text{Var}_D(B)$$
Adjusted variance

- We partition the variance matrix of $B$ into two variance components:

$$\text{Var}(B) = \text{Var}(E_D(B)) + \text{Var}(A_B(D))$$
$$= \text{RVar}_D(B) + \text{Var}_D(B)$$

- These are the resolved variance matrix and the adjusted variance matrix (i.e. explained and residual variation).
Adjusted variance

- We partition the variance matrix of $B$ into two variance components:

$$ \text{Var}(B) = \text{Var}(E_D(B)) + \text{Var}(A_B(D)) $$

$$ = R\text{Var}_D(B) + \text{Var}_D(B) $$

- These are the resolved variance matrix and the adjusted variance matrix (i.e. explained and residual variation).

- The variance matrices are calculated as

$$ \text{Var}_D(B) = \text{Var}(B) - \text{Cov}(B, D) \text{Var}(D)^\dagger \text{Cov}(D, B) , $$

$$ R\text{Var}_D(B) = \text{Cov}(B, D) \text{Var}(D)^\dagger \text{Cov}(D, B) . $$
Adjusted variance

- We partition the variance matrix of $B$ into two variance components:

$$\text{Var}(B) = \text{Var}(E_D(B)) + \text{Var}(A_B(D))$$

$$= \text{RVar}_D(B) + \text{Var}_D(B)$$

- These are the resolved variance matrix and the adjusted variance matrix (i.e. explained and residual variation).

- The variance matrices are calculated as

$$\text{Var}_D(B) = \text{Var}(B) - \text{Cov}(B, D)\text{Var}(D)^\dagger\text{Cov}(D, B),$$

$$\text{RVar}_D(B) = \text{Cov}(B, D)\text{Var}(D)^\dagger\text{Cov}(D, B).$$

- Our variance matrices must be non-negative definite.
Adjusted variance

- We partition the variance matrix of $B$ into two variance components:

$$\text{Var}(B) = \text{Var}(E_D(B)) + \text{Var}(A_B(D))$$

$$= \text{RVar}_D(B) + \text{Var}_D(B)$$

- These are the resolved variance matrix and the adjusted variance matrix (i.e. explained and residual variation).

- The variance matrices are calculated as

$$\text{Var}_D(B) = \text{Var}(B) - \text{Cov}(B, D) \text{Var}(D)^\dagger \text{Cov}(D, B),$$

$$\text{RVar}_D(B) = \text{Cov}(B, D) \text{Var}(D)^\dagger \text{Cov}(D, B).$$

- Our variance matrices must be non-negative definite.
- We use the Moore-Penrose generalized inverse ($A^\dagger$) to allow for degeneracy.
Resolution

We summarize the expected effect of the data $D$ for the adjustment of $B$ by a scale-free measure which we call the resolution of $B$ induced by $D$,

$$R_D(B) = 1 - \frac{\text{Var}_D(B)}{\text{Var}(B)} = \frac{\text{Var}(E_D(B))}{\text{Var}(B)}.$$

The resolution lies between 0 and 1, and in general, small (large) resolutions imply that the information has little (much) linear predictive value, given the prior specification.
Resolution

- We summarize the expected effect of the data $D$ for the adjustment of $B$ by a scale-free measure which we call the resolution of $B$ induced by $D$,

$$R_D(B) = 1 - \frac{\text{Var}_D(B)}{\text{Var}(B)} = \frac{\text{Var}(E_D(B))}{\text{Var}(B)}.$$ 

- The resolution lies between 0 and 1, and in general, small (large) resolutions imply that the information has little (much) linear predictive value, given the prior specification.
Resolution

- We summarize the expected effect of the data $D$ for the adjustment of $B$ by a scale-free measure which we call the resolution of $B$ induced by $D$,

$$R_D(B) = 1 - \frac{\text{Var}_D(B)}{\text{Var}(B)} = \frac{\text{Var}(E_D(B))}{\text{Var}(B)}.$$ 

- The resolution lies between 0 and 1, and in general, small (large) resolutions imply that the information has little (much) linear predictive value, given the prior specification.

- Similar in spirit to an $R^2$ measure for the adjustment.
Example: The Adjustment

We can calculate our adjusted expectations for points $B$ given $D$ algebraically as:

$$E_D(B_1) = 0.381D_1 + 0.507D_2 + 0$$
$$E_D(B_2) = 0.051D_1 + 0.961D_2 + 0$$
Example: The Adjustment

- We can calculate our adjusted expectations for points $B$ given $D$ algebraically as:

$$E_D(B_1) = 0.381D_1 + 0.507D_2 + 0$$

$$E_D(B_2) = 0.051D_1 + 0.961D_2 + 0$$

- We see that $B_2$ is mainly determined by the value of $D_2$ — unsurprising given the strength of $\text{Corr}(B_2, D_2)$. 

Example: The Adjustment

- We can calculate our adjusted expectations for points $B$ given $D$ algebraically as:

  \[ E_D(B_1) = 0.381D_1 + 0.507D_2 + 0 \]
  \[ E_D(B_2) = 0.051D_1 + 0.961D_2 + 0 \]

- We see that $B_2$ is mainly determined by the value of $D_2$ – unsurprising given the strength of $\text{Corr}(B_2, D_2)$.

- We can also calculate the adjusted variance and resolutions

  \[ \text{Var}_D(B) = \begin{pmatrix} 49.06 & -5.83 \\ -5.83 & 4.64 \end{pmatrix}, \quad \text{R}_D(B) = \begin{pmatrix} 0.509 \\ 0.954 \end{pmatrix} \]

- We can see that we resolve much of the uncertainty about $B_2$
Example: Variance Partition

We can decompose the prior variance into its resolved and unresolved portions:

$$\text{Var}(B) = \text{RVar}_D(B) + \text{Var}_D(B)$$

\[
\begin{pmatrix}
100.00 & 55.71 \\
55.71 & 100
\end{pmatrix}
= \begin{pmatrix}
50.94 & 61.54 \\
61.54 & 95.36
\end{pmatrix}
+ \begin{pmatrix}
49.06 & -5.83 \\
-5.83 & 4.64
\end{pmatrix}
\]
The observed adjustment

- Given the observed value $d$ of $D$, we can calculate the observed adjusted expectation

$$E_d(B) = E(B) + \text{Cov}(B, D) \text{Var}(D)^\dagger(d - E(D)).$$
The observed adjustment

- Given the observed value \( d \) of \( D \), we can calculate the observed adjusted expectation

\[
E_d(B) = E(B) + \text{Cov}(B, D) \text{Var}(D)^\dagger(d - E(D)).
\]

- For our example, we observe \( d = (-8, 10) \) and the corresponding observed adjusted expectations are:

\[
E_d(B) = \begin{pmatrix} 2.02 \\ 9.20 \end{pmatrix}
\]

- Having observed \( D = d \), we notice that our adjusted expectations have both increased
The observed adjustment

- Given the observed value \( d \) of \( D \), we can calculate the observed adjusted expectation

\[
E_d(B) = E(B) + \text{Cov}(B, D) \text{Var}(D)^\dagger (d - E(D)).
\]

- For our example, we observe \( d = (-8, 10) \) and the corresponding observed adjusted expectations are:

\[
E_d(B) = \begin{pmatrix} 2.02 \\ 9.20 \end{pmatrix}
\]

- Having observed \( D = d \), we notice that our adjusted expectations have both increased

- \( B_1 \) is weakly correlated with \( D \) and so is adjusted only a little, whereas \( B_2 \) is strongly correlated to \( D_2 \) and so its expectation shifts substantially towards the value \( d_2 = 10 \).
Interpretation
Interpretations of belief adjustment

- **An approximation**
  - If we’re fully Bayesian, then adjusted expectation is a tractable linear approximation to the full Bayes conditional expectation.
  - Adjusted variance is then an easily-computable upper bound on the full Bayes preposterior risk, under quadratic loss.
Interpretations of belief adjustment

- **An approximation**
  - If we’re fully Bayesian, then adjusted expectation is a tractable linear approximation to the full Bayes conditional expectation.
  - Adjusted variance is then an easily-computable upper bound on the full Bayes preposterior risk, under quadratic loss.

- **An estimator**
  - $E_D(B)$ is an ‘estimator’ of the value of $B$, which combines the data with simple aspects of our prior beliefs in a plausible manner.
  - Adjusted variance is then the mean-squared error of the estimator $E_D(B)$. 
Interpretations of belief adjustment

- **An approximation**
  - If we’re fully Bayesian, then adjusted expectation is a tractable linear approximation to the full Bayes conditional expectation.
  - Adjusted variance is then an easily-computable upper bound on the full Bayes preposterior risk, under quadratic loss.

- **An estimator**
  - $E_D(B)$ is an ‘estimator’ of the value of $B$, which combines the data with simple aspects of our prior beliefs in a plausible manner.
  - Adjusted variance is then the mean-squared error of the estimator $E_D(B)$.

- **A primitive**
  - Adjusted expectation is a primitive quantification of further aspects of our beliefs about $B$ having ‘accounted for’ $D$.
  - Adjusted variance is also a primitive, but applied to the ‘residual variance’ in $B$ having removed the effects of $D$. 

Jonathan Cumming, Ian Vernon
Introduction to Bayes Linear Statistics
Conditional Expectation

- The conditional expectation of $B|D$ is the value you would specify under the penalty

$$L_C = \sum_i cD_i [B - E(B|D_i)]^2$$
Conditional Expectation

- The conditional expectation of $B|D$ is the value you would specify under the penalty $L_C = \sum_i cD_i [B - E(B|D_i)]^2$

- If $D$ is a partition, so $D_i \in \{0, 1\}$ and $\sum_i D_i = 1$, then the adjusted expectation minimises $L_A = \sum_i cD_i [B - x_i]^2$. So we choose $x_i$ to be the conditional expectation, and

$$E_D(B) = \sum_i E(B|D_i)D_i$$
Conditional Expectation

The conditional expectation of $B|D$ is the value you would specify under the penalty $L_C = \sum_i cD_i[B - E(B|D_i)]^2$.

If $D$ is a partition, so $D_i \in \{0, 1\}$ and $\sum_i D_i = 1$, then the adjusted expectation minimises $L_A = \sum_i cD_i[B - x_i]^2$. So we choose $x_i$ to be the conditional expectation, and

$$E_D(B) = \sum_i E(B|D_i)D_i$$

So when $D$ is a partition, the adjusted and conditional expectations are identical.
Conditional Expectation

- The conditional expectation of $B|D$ is the value you would specify under the penalty $L_C = \sum_i cD_i[B - E(B|D_i)]^2$

- If $D$ is a partition, so $D_i \in \{0, 1\}$ and $\sum_i D_i = 1$, then the adjusted expectation minimises $L_A = \sum_i cD_i[B - x_i]^2$. So we choose $x_i$ to be the conditional expectation, and

$$E_D(B) = \sum_i E(B|D_i)D_i$$

- So when $D$ is a partition, the adjusted and conditional expectations are identical

- Adjusted expectation does not require $D$ to be a partition, and so can be considered as a generalization of conditional expectation
Extension to linear combinations

Let \( \langle B \rangle \) be the set of all linear combinations of \( B \)
Extension to linear combinations

- Let $\langle B \rangle$ be the set of all linear combinations of $B$
- If $X = h^T B \in \langle B \rangle$, then we can write

$$E(X) = h^T E(B), \ Var(X) = h^T Var(B) h.$$
Extension to linear combinations

- Let $\langle B \rangle$ be the set of all linear combinations of $B$
- If $X = h^T B \in \langle B \rangle$, then we can write

$$E(X) = h^T E(B), \quad \text{Var}(X) = h^T \text{Var}(B) h.$$ 

- So by specifying $E(B)$ and $\text{Var}(B)$ we have \textit{implicitly specified} expectations and variances for all elements of $\langle B \rangle$
Extension to linear combinations

- Let $\langle B \rangle$ be the set of all linear combinations of $B$
- If $X = h^T B \in \langle B \rangle$, then we can write
  \[ E(X) = h^T E(B), \quad \text{Var}(X) = h^T \text{Var}(B) h. \]
- So by specifying $E(B)$ and $\text{Var}(B)$ we have implicitly specified expectations and variances for all elements of $\langle B \rangle$
- Similarly, by calculating $E_D(B)$ and $\text{Var}_D(B)$, we have implicitly calculated the adjustment for all $X \in \langle B \rangle$
Diagnostics
Data and Diagnostics

Once data has been observed (first for $D$ and then for $B$) we can perform diagnostics.
Data and Diagnostics

- Once data has been observed (first for $D$ and then for $B$) we can perform diagnostics.
- The Bayes linear methodology has a rich variety of diagnostic tools available (more than in a fully Bayesian analysis).
Data and Diagnostics

- Once data has been observed (first for $D$ and then for $B$) we can perform diagnostics.
- The Bayes linear methodology has a rich variety of diagnostic tools available (more than in a fully Bayesian analysis).
- We can perform diagnostics on individual random quantities, or on collections of random quantities.
Data and Diagnostics

- Once data has been observed (first for $D$ and then for $B$) we can perform diagnostics.
- The Bayes linear methodology has a rich variety of diagnostic tools available (more than in a fully Bayesian analysis).
- We can perform diagnostics on individual random quantities, or on collections of random quantities.
- Three important versions are:
  - Prior Diagnostics.
  - Adjustment Diagnostics.
  - Final Observation Diagnostics.
Each prior belief statement that we make describes our beliefs about some random quantity.
Prior Diagnostics

- Each prior belief statement that we make describes our beliefs about some random quantity.
- If we observe that quantity, we may compare what we expect to happen with what actually happens.

\[ S(d_i) = \frac{d_i - E(D_i)}{\sqrt{Var(D_i)}} \]
\[ Dis(d) = \left( d_i - E(D_i) \right)^2 Var(D_i) = S(d_i)^2 \]

\[ E(S(d_i)) = 0 \text{ and } Var(S(d_i)) = 1 \text{, so if we observe } S(d_i) \text{ greater than about 3 this suggests an inconsistency.} \]
Prior Diagnostics

- Each prior belief statement that we make describes our beliefs about some random quantity.
- If we observe that quantity, we may compare what we expect to happen with what actually happens.
- Once we observe the values of $D = d$, we can check whether the data is consistent with our prior specifications.
Prior Diagnostics

- Each prior belief statement that we make describes our beliefs about some random quantity.
- If we observe that quantity, we may compare what we expect to happen with what actually happens.
- Once we observe the values of $D = d$, we can check whether the data is consistent with our prior specifications.
- For a single random quantity, we can calculate the standardized change and the discrepancy:

$$S(d_i) = \frac{d_i - E(D_i)}{\sqrt{Var(D_i)}}, \quad \text{Dis}(d) = \frac{[d_i - E(D_i)]^2}{Var(D_i)} = S(d_i)^2$$
Each prior belief statement that we make describes our beliefs about some random quantity.

If we observe that quantity, we may compare what we expect to happen with what actually happens.

Once we observe the values of $D = d$, we can check whether the data is consistent with our prior specifications.

For a single random quantity, we can calculate the standardized change and the discrepancy:

$$S(d_i) = \frac{d_i - E(D_i)}{\sqrt{\text{Var}(D_i)}}$$
$$\text{Dis}(d) = \frac{(d_i - E(D_i))^2}{\text{Var}(D_i)} = S(d_i)^2$$

$E(S(d_i)) = 0$ and $\text{Var}(S(d_i)) = 1$, so if we observe $S(d_i)$ greater than about 3 this suggests an inconsistency.
Discrepancy Ratio

For the entire collection, the natural counterpart of the discrepancy is the Mahalanobis distance:

\[ \text{Dis}(d) = (d - \text{E}(D))^T \text{Var}(D)^\dagger (d - \text{E}(D)). \]
Discrepancy Ratio

- For the entire collection, the natural counterpart of the discrepancy is the Mahalanobis distance:
  \[
  \text{Dis}(d) = (d - E(D))^T \text{Var}(D)^\dagger (d - E(D)).
  \]

- The prior expected value of \( \text{Dis}(d) \) is given by
  \[
  E(\text{Dis}(d)) = \text{rk}\{\text{Var}(D)\}.
  \]
Discrepancy Ratio

- For the entire collection, the natural counterpart of the discrepancy is the Mahalanobis distance:
  \[
  \text{Dis}(d) = (d - \mathbb{E}(D))^T \text{Var}(D)^\dagger (d - \mathbb{E}(D)).
  \]
- The prior expected value of \(\text{Dis}(d)\) is given by
  \[
  \mathbb{E}(\text{Dis}(d)) = \text{rk}\{\text{Var}(D)\}
  \]
- NB: if we pretend \(D\) is Normal, then \(\text{Dis}(d)\) would be \(\chi^2\)
Discrepancy Ratio

- For the entire collection, the natural counterpart of the discrepancy is the Mahalanobis distance:

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

- The prior expected value of \text{Dis}(d) is given by

\[
\text{E}(\text{Dis}(d)) = \text{rk}\{\text{Var}(D)\}
\]

- NB: if we pretend \( D \) is Normal, then \text{Dis}(d) would be \( \chi^2 \)

- We can then normalise the discrepancy, to obtain the discrepancy ratio for \( d \)

\[
\text{Dr}(d) = \frac{\text{Dis}(d)}{\text{rk}\{\text{Var}(D)\}},
\]

which has prior expectation \( \text{E}(\text{Dr}(d)) = 1 \).
Discrepancy Ratio

- For the entire collection, the natural counterpart of the discrepancy is the Mahalanobis distance:
  \[
  \text{Dis}(d) = (d - \mathbb{E}(D))^T \text{Var}(D)^+ (d - \mathbb{E}(D)).
  \]

- The prior expected value of \( \text{Dis}(d) \) is given by
  \[
  \mathbb{E}(\text{Dis}(d)) = \text{rk}\{\text{Var}(D)\}
  \]

- NB: if we pretend \( D \) is Normal, then \( \text{Dis}(d) \) would be \( \chi^2 \)

- We can then normalise the discrepancy, to obtain the discrepancy ratio for \( d \)
  \[
  \text{Dr}(d) = \frac{\text{Dis}(d)}{\text{rk}\{\text{Var}(D)\}},
  \]
  which has prior expectation \( \mathbb{E}(\text{Dr}(d)) = 1 \).

- Large \( \text{Dr}(d) \) will of course also suggest inconsistencies.
Further Topics
Canonical Analysis
Our belief specification for $B$ and our adjustment by $D$ implies specifications and adjustments for all linear combinations in $\langle B \rangle$.
Canonical analysis

- Our belief specification for $B$ and our adjustment by $D$ implies specifications and adjustments for all linear combinations in $\langle B \rangle$.
- We can explore the (possibly complex) changes in beliefs about $\langle B \rangle$ induced by the adjustment via a canonical analysis.
Canonical analysis

- Our belief specification for $B$ and our adjustment by $D$ implies specifications and adjustments for all linear combinations in $\langle B \rangle$.
- We can explore the (possibly complex) changes in beliefs about $\langle B \rangle$ induced by the adjustment via a canonical analysis.
- A key component of the canonical analysis is the resolution transform matrix defined as

$$T_{B:D} = \text{Var}(B)^\dagger \text{Cov}(B, D) \text{Var}(D)^\dagger \text{Cov}(D, B).$$
Canonical analysis

- Our belief specification for $B$ and our adjustment by $D$ implies specifications and adjustments for all linear combinations in $\langle B \rangle$.
- We can explore the (possibly complex) changes in beliefs about $\langle B \rangle$ induced by the adjustment via a canonical analysis.
- A key component of the canonical analysis is the resolution transform matrix defined as

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

- $\mathbb{T}_{B:D}$ has the property that $\text{Var}(B)\mathbb{T}_{B:D} = R\text{Var}_D(B)$.
Our belief specification for $B$ and our adjustment by $D$ implies specifications and adjustments for all linear combinations in $\langle B \rangle$.

We can explore the (possibly complex) changes in beliefs about $\langle B \rangle$ induced by the adjustment via a canonical analysis.

A key component of the canonical analysis is the resolution transform matrix defined as

$$T_{B:D} = \Var(B)^\dagger \Cov(B, D) \Var(D)^\dagger \Cov(D, B).$$

$T_{B:D}$ has the property that $\Var(B)T_{B:D} = \RVar_D(B)$

The eigenstructure of $T_{B:D}$ summarises all the effects of belief adjustment.
Canonical analysis

- Our belief specification for $B$ and our adjustment by $D$ implies specifications and adjustments for all linear combinations in $\langle B \rangle$.
- We can explore the (possibly complex) changes in beliefs about $\langle B \rangle$ induced by the adjustment via a canonical analysis.
- A key component of the canonical analysis is the resolution transform matrix defined as

$$T_{B:D} = \text{Var}(B)^\dagger \text{Cov}(B, D) \text{Var}(D)^\dagger \text{Cov}(D, B).$$

$T_{B:D}$ has the property that $\text{Var}(B)T_{B:D} = \text{RVar}_D(B)$.

- The eigenstructure of $T_{B:D}$ summarises all the effects of belief adjustment.
- Let the normed right eigenvectors of $T_{B:D}$ be $v_1, \ldots, v_{r_B}$, ordered by eigenvalues $1 \geq \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{r_B} \geq 0$ and scaled as $v_i^T \text{Var}(B)v_i = 1$. 

Jonathan Cumming, Ian Vernon
Introduction to Bayes Linear Statistics
Canonical directions

- We define the $i$th canonical direction as

$$Y_i = v_i^T (B - E(B))$$
Canonical directions

- We define the $i$th canonical direction as
  \[ Y_i = v_i^T (B - E(B)) \]
- The canonical directions have the following properties
  \[ E(Y_i) = 0, \quad \text{Var}(Y_i) = 1, \quad \text{Corr}(Y_i, Y_j) = 0 \]
  \[ \text{RVar}_D(Y_i) = \lambda_i, \quad \text{Var}_D(Y_i) = 1 - \lambda_i, \]
Canonical directions

- We define the $i$th canonical direction as
  \[ Y_i = v_i^T (B - \mathbb{E}(B)) \]
- The canonical directions have the following properties
  \[
  \mathbb{E}(Y_i) = 0, \quad \text{Var}(Y_i) = 1, \quad \text{Corr}(Y_i, Y_j) = 0
  \]
  \[
  \text{RVar}_D(Y_i) = \lambda_i, \quad \text{Var}_D(Y_i) = 1 - \lambda_i,
  \]
- So the collection \{ $Y_1, Y_2, \ldots$ \} forms a mutually uncorrelated ‘grid’ of directions over $\langle B \rangle$, summarizing the effects of the adjustment.
Canonical directions

- We define the $i$th canonical direction as

$$Y_i = v_i^T (B - E(B))$$

- The canonical directions have the following properties

$$E(Y_i) = 0, \quad \text{Var}(Y_i) = 1, \quad \text{Corr}(Y_i, Y_j) = 0$$

$$\text{RVar}_D(Y_i) = \lambda_i, \quad \text{Var}_D(Y_i) = 1 - \lambda_i,$$

- So the collection $\{Y_1, Y_2, \ldots\}$ forms a mutually uncorrelated ‘grid’ of directions over $\langle B \rangle$, summarizing the effects of the adjustment.

- $Y_1$ is the quantity we learn most about. $Y_2$ is the quantity we learn next most about, given that it is uncorrelated with $Y_1$.

- $Y_{rk\{B\}}$ is the quantity we learn least about.

- Relationship to canonical correlation analysis (and PCA)
Canonical properties and system resolution

- Each $X \in \langle B \rangle$ can be expressed using the canonical structure as

$$X - \text{E}(X) = \sum_i \text{Cov}(X, Y_i) Y_i,$$

and $\text{RVar}_D(X) = \sum_i \lambda_i (\text{Corr}(X, Y_i))^2$
Canonical properties and system resolution

- Each $X \in \langle B \rangle$ can be expressed using the canonical structure as

$$X - \mathbb{E}(X) = \sum_i \text{Cov}(X, Y_i) Y_i,$$

and $\text{RVar}_D(X) = \sum_i \lambda_i (\text{Corr}(X, Y_i))^2$

- We can use this structure to express the resolved uncertainty for the entire collection $\langle B \rangle$ given adjustment by $D$ via the resolved uncertainty and the system resolution

$$\text{RU}_D(B) = \sum_i \lambda_i, \quad \text{R}_D(B) = \frac{1}{\text{rk}\{B\}} \sum_i \lambda_i$$
Canonical properties and system resolution

- Each $X \in \langle B \rangle$ can be expressed using the canonical structure as

\[ X - E(X) = \sum_i \text{Cov}(X, Y_i) Y_i, \]

and $R\text{Var}_D(X) = \sum_i \lambda_i (\text{Corr}(X, Y_i))^2$

- We can use this structure to express the resolved uncertainty for the entire collection $\langle B \rangle$ given adjustment by $D$ via the resolved uncertainty and the system resolution

\[ R\text{U}_D(B) = \sum_i \lambda_i, \quad R_D(B) = \frac{1}{\text{rk}\{B\}} \sum_i \lambda_i \]

- $R_D(B)$ is a scalar summary of the effectiveness of the adjustment by $D$ for the entire collection $\langle B \rangle$
Partial Analysis
Partial Analysis

- Suppose we have already adjusted out beliefs about $B$ given data, $D$
  - Now suppose we get even more data $F$, how should we further adjust our beliefs about $B$?
Partial Analysis

- Suppose we have already adjusted out beliefs about $B$ given data, $D$
  - Now suppose we get even more data $F$, how should we further adjust our beliefs about $B$?
- Suppose we have already adjusted our beliefs about $B$ given data, $H = D \cup F$
  - What were the individual effects of adjusting by $D$ or $F$?
Partial Analysis

- Suppose we have already adjusted out beliefs about $B$ given data, $D$
  - Now suppose we get even more data $F$, how should we further adjust our beliefs about $B$?
- Suppose we have already adjusted our beliefs about $B$ given data, $H = D \cup F$
  - What were the individual effects of adjusting by $D$ or $F$?
- To answer either of these questions would require a partial analysis, where we consider the effects of subsets of the data on our beliefs
Partial adjustments

- If we adjust beliefs sequentially, then we can separate and scrutinize the adjustments at each stage.
Partial Analysis

Partial adjustments

- If we adjust beliefs sequentially, then we can separate and scrutinize the adjustments at each stage.
- We evaluate partial adjustments which represent the change in adjustment as we accumulate data.
Partial adjustments

- If we adjust beliefs sequentially, then we can separate and scrutinize the adjustments at each stage.
- We evaluate **partial adjustments** which represent the change in adjustment as we accumulate data.
- Suppose we intend to adjust our beliefs about $B$ by observations on $D$ and $F$, we adjust $B$ by $(D \cup F)$ but separate the effects of the subsets by adjusting $B$ in stages, first by $D$, then adding $F$ (or vice versa).
Partial adjustments

- If we adjust beliefs sequentially, then we can separate and scrutinize the adjustments at each stage.
- We evaluate partial adjustments which represent the change in adjustment as we accumulate data.
- Suppose we intend to adjust our beliefs about $B$ by observations on $D$ and $F$, we adjust $B$ by $(D \cup F)$ but separate the effects of the subsets by adjusting $B$ in stages, first by $D$, then adding $F$ (or vice versa).
- How do we separate the effects of $D$ and $F$ on $B$?
Partial Analysis

Separating things out

If $D \perp F$, then adjusted expectations are additive so

$$E_{D \cup F}(B - E(B)) = E_D(B - E(B)) + E_F(B - E(B))$$
### Separating things out

- If $D \perp F$, then adjusted expectations are additive so
  \[
  E_{D\cup F}(B - E(B)) = E_D(B - E(B)) + E_F(B - E(B))
  \]
- If $D$ and $F$ are correlated, then we obtain a similar expression by removing the ‘common variability’ between $F$ and $D$. 
Partial Analysis

Separating things out

- If $D \perp F$, then adjusted expectations are additive so

$$E_{D \cup F}(B - E(B)) = E_D(B - E(B)) + E_F(B - E(B))$$

- If $D$ and $F$ are correlated, then we obtain a similar expression by removing the ‘common variability’ between $F$ and $D$.

- For any $D$, $F$, the vectors $D$ and $\mathbb{A}_F(D) = F - E_D(F)$ are uncorrelated.

- So, for any $D$, $F$

$$E_{D \cup F}(B - E(B)) = E_D(B - E(B)) + E_{\mathbb{A}_F(D)}(B - E(B))$$
The partial adjustment

- The partial adjustment of $B$ by $F$ given $D$, denoted $E_{[F/D]}(B)$, is

$$E_{[F/D]}(B) = E_{D\cup F}(B) - E_D(B) = E_{A_F(D)}(B - E(B))$$
The partial adjustment

- The partial adjustment of $B$ by $F$ given $D$, denoted $E_{[F/D]}(B)$, is

$$E_{[F/D]}(B) = E_{D∪F}(B) - E_D(B) = E_{AF(D)}(B - E(B))$$

- We can partition the variance in several ways

$$\text{Var}(B) = R\text{Var}_D(B) + \text{Var}_D(B)$$
$$= R\text{Var}_D(B) + R\text{Var}_{[F/D]}(B) + \text{Var}_{D∪F}(B)$$
$$= R\text{Var}_{D∪F}(B) + \text{Var}_{D∪F}(B)$$
The partial adjustment

- The partial adjustment of $B$ by $F$ given $D$, denoted $E_{[F/D]}(B)$, is

$$E_{[F/D]}(B) = E_{D∪F}(B) - E_D(B) = E_{AF}(D)(B - E(B))$$

- We can partition the variance in several ways

$$\text{Var}(B) = \text{RVar}_D(B) + \text{Var}_D(B)$$
$$= \text{RVar}_D(B) + \text{RVar}_{[F/D]}(B) + \text{Var}_{D∪F}(B)$$
$$= \text{RVar}_{D∪F}(B) + \text{Var}_{D∪F}(B)$$

- The partial resolved variance matrix of $B$ by $F$ given $D$ is

$$\text{RVar}_{[F/D]}(B) = \text{Var}(E_{[F/D]}(B))$$

Jonathan Cumming, Ian Vernon
Introduction to Bayes Linear Statistics
The end
We have seen:

- How we represent our beliefs – using expectation as primitive
- How we would update our beliefs – the BL adjustment
- How we can investigate potential problems in our belief specification – diagnostics
- How we can understand how our beliefs are affected by the data – canonical analysis
- How we would incorporate additional information – partial analysis