Bayes linear revision for plates

David Wooff

Durham University

LMS Durham Symposium
Mathematical Aspects of Graphical Models

Bayes linear methods: whirlwind tour

- Statistical methods based on expectation and variance-covariance. No probability distributions!
- The **adjusted expectation** for collection B given collection D is

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

- The **adjusted version** of the B given D is the 'residual' vector

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

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

$$B = \mathrm{E}_D(B) + \mathbb{A}_D(B),$$

- We partition the variance matrix of B into 2 variance components

$$\mathrm{Var}(B) = \mathrm{Var}(\mathrm{E}_D(B)) + \mathrm{Var}(\mathbb{A}_D(B))$$

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

Variance objects

- The variance matrices are calculated as

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

- Our variance matrices must be non-negative definite with $0 < \mathbf{tr}\{\cdot\} < \infty$. So, they might be singular but must reflect at least one linear combination with positive variance. That is, $0 \leq a^T \Sigma a < \infty \; \forall a$, and $a^T \Sigma a > 0$ for some a.
- We use the Moore-Penrose generalized inverse (allows for degeneracy).
- The **adjusted covariance matrix** and **resolved covariance matrix** are defined similarly:

$$\mathrm{Cov}_D(B_1, B_2) = \mathrm{Cov}(B_1, B_2) - \mathrm{Cov}(B_1, D)\mathrm{Var}(D)^\dagger \mathrm{Cov}(D, B_2),$$
$$\mathrm{RCov}_D(B_1, B_2) = \mathrm{Cov}(B_1, D)\mathrm{Var}(D)^\dagger \mathrm{Cov}(D, B_2).$$

- Similarities to full Bayesian updating for Gaussian quantities, and links to classical statistical methodology, e.g. canonical correlation analysis.

Data and diagnostics

- Bayes linear methods offer a number of diagnostics, mostly comparing actual to expected behaviour, and most based on Mahalanobis distances of various kinds.
- For example, a measure of the relative difference between the data d and their prior expectations $\mathrm{E}(D)$, is the **discrepancy**, $\mathrm{Dis}(d)$, computed as 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, $\mathrm{E}(\mathrm{Dis}(D)) = \mathbf{rk}\{\mathrm{Var}(D)\}$. , and so we my define the **discrepancy ratio**

$$\mathrm{Dr}(d) = \frac{\mathrm{Dis}(d)}{\mathbf{rk}\{\mathrm{Var}(D)\}},$$

  as 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 may derive similar diagnostics for the observed adjusted expectations, defining the **adjustment discrepancy** as:

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

- This is essentially the squared change in expectation from prior to posterior, relative to variance explained, and can be compared to its expected value, $\mathbf{rk}\{\mathrm{RVar}_d(B)\}$.

Canonical structure and the resolution transform

- We are careful to pay attention to all possible linear combinations of our random quantities for the analysis of beliefs.
- The **canonical structure** gives a natural framework underpinning such analysis.
- The **resolution transform matrix** is defined as

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

- We can calculate the canonical directions $Z_1, \ldots, Z_{r_B}$ by finding the normed right eigenvectors of $\mathbb{T}_{B:D}$, which we write $v_1, \ldots, v_{r_B}$, ordered by eigenvalues

$$1 \geq \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{r_B} \geq 0$$

and scaled, for each i, as $v_i^T \mathrm{Var}(B) v_i = 1$, so that

$$Y_i = v_i^T(B - \mathrm{E}(B)), \quad \text{and} \quad \mathrm{Var}_D(Y_i) = 1 - \lambda_i.$$

Canonical structure and the resolution transform II

- The resolution transform is of intrinsic interest as the object which summarizes, through the eigenstructure, all of the effects of the belief adjustment.
- There is a strong relationship between this transform and classical canonical correlation analysis.
- The collection $\{Z_1, Z_2, \ldots\}$ forms a mutually uncorrelated 'grid' of directions over $\langle B \rangle$, summarizing the effects of the adjustment.
- $\mathrm{E}(Z_i) = 0$, $\mathrm{Var}(Z_i) = 1$, and $\mathrm{Var}_D(Z_i) = 1 - \lambda_i$. So, these constructed random quantities are uncorrelated with prior expectation zero and prior variance one. Linear fitting on D is expected to reduce uncertainty from one to $1 - \lambda_i$, so that λ_i is the proportion of variance resolved, $\mathrm{R}_D(Z_i) = \lambda_i$.
- Z_1 is the quantity we learn most about. Z_2 is the quantity we learn next most about, given that it is uncorrelated with Z_1. $Z_{\mathbf{rk}\{B\}}$ is the quantity we learn least about.

Canonical structure and the resolution transform III

- Each quantity, $X = a^T B$, may be resolved along the canonical directions as

$$X - \mathrm{E}(X) = \sum_i \mathrm{Cov}(X, Z_i) Z_i.$$

- Thus, all changes in expectation may be deduced via the canonical structure.
- The proportion of variance explained can be determined as a linear combination of the eigenvalues of the resolution transform matrix, and is bounded by the largest and smallest eigenvalues.
- The adjustment of belief can similarly be carried out on this grid, e.g.

$$\mathrm{E}_D(X) = \sum_i \mathrm{Cov}(X, Z_i) \mathrm{E}_D(Z_i).$$

The bearing

- The **bearing for the adjustment** of B by $D = d$ is

$$\mathbb{Z}_d(B) = [\mathrm{E}_d(B) - \mathrm{E}(B)]^T \mathrm{Var}(B)^\dagger [B - \mathrm{E}(B)].$$

- It expresses both the direction and the magnitude of the change between prior and adjusted beliefs, relative to the prior covariance specification because, for any $F = u^T B \in \langle B \rangle$,

$$\mathrm{Cov}(F, \mathbb{Z}_d(B)) = \mathrm{E}_d(F) - \mathrm{E}(F).$$

- For any X uncorrelated with $\mathbb{Z}_d(B)$ we have $\mathrm{E}_d(X) = \mathrm{E}(X)$ (no change in expectation).
- The biggest possible expected squared change in expectation, relative to prior variance, is for the linear combination given by $\mathbb{Z}_d(B)$, and the amount of change is defined to be the **size of the adjustment**:

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

- A natural diagnostic for assessing the magnitude of an adjustment is to compare the largest standardized change in expectation that we observe to our expectation for the magnitude of the largest change, evaluated prior to observing D. We can show that $\mathrm{E}(\mathrm{Size}_D(B)) = \sum \lambda_i$ is the explained uncertainty over the structure.
- The **size ratio for the adjustment of B by D** is

$$\mathrm{Sr}_d(B) = [\mathrm{E}_d(B) - \mathrm{E}(B)]^T \mathrm{Var}(B)^\dagger [\mathrm{E}_d(B) - \mathrm{E}(B)] / \sum \lambda_i.$$

- Large ($>> 1$) or small ($<< 1$) values suggest contradictions in behaviour.

Partial adjustments

- We can carry out belief adjustments sequentially. This is informative when there is a natural time ordering or when we want to scrutinize the process.
- Each change in adjustment is a **partial** adjustment. Every summary and diagnostic which we have discussed can be calculated for the partial adjustment, typically relating to the prior variance resolved by the partial fit. Hence, there is a **partial resolution transform matrix**, a **partial bearing** and a **partial size ratio**, all of which we routinely inspect.
- When we adjust beliefs in stages, e.g. B by D and then F, the expected sizes of the respective adjustments are additive:

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

- However, the observed sizes of the adjustments are not additive. The size of each adjustment is the variance of the corresponding bearing:

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

- The observed value of the covariance term $\mathrm{Cov}(\mathbb{Z}_{d}(B), \mathbb{Z}_{[f/d]}(B))$ may be taken to expresses the degree of support or conflict between the two collections of evidence in determining the revision of beliefs.

Path correlation diagnostics

- As a summary, we define the **path correlation** to be

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

  the correlation between the initial and partial bearings.
- If the path correlation is near $+1$ then the size of the adjustment of B by $D \cup F$ is much larger than the sum of the size of the adjustment by D and the size of the partial adjustment by $[F/D]$; we view the two collections of data as complementary: their combined effect in changing our beliefs is greater than the sum of the individual effects.
- If the path correlation is near -1 then the two collections are giving 'contradictory' messages which give smaller overall changes in belief, in combination, than we would expect from the individual adjustments with D and $[F/D]$; e.g. each of the individual changes in belief might be surprisingly large but the overall change in belief might be small, masking these differences.
- The importance of such conflict depends on the magnitudes of the various changes in beliefs, but we do wish to distinguish between analyses with a sequence of small changes in expectation, and analyses where individual data sources suggested large changes in beliefs but these were of a contradictory nature and so cancelled out each other.

Regression example

- Taken from Box & Tiao (1973), analysed in Goldstein & Wooff (2007).
- A chemical process leads to a product Y and a by-product Z. Yields of both products are thought to be related to the temperature of the process, X. Twelve experiments are performed with different temperature settings (degrees Fahrenheit) to study the effect of temperature on yield.

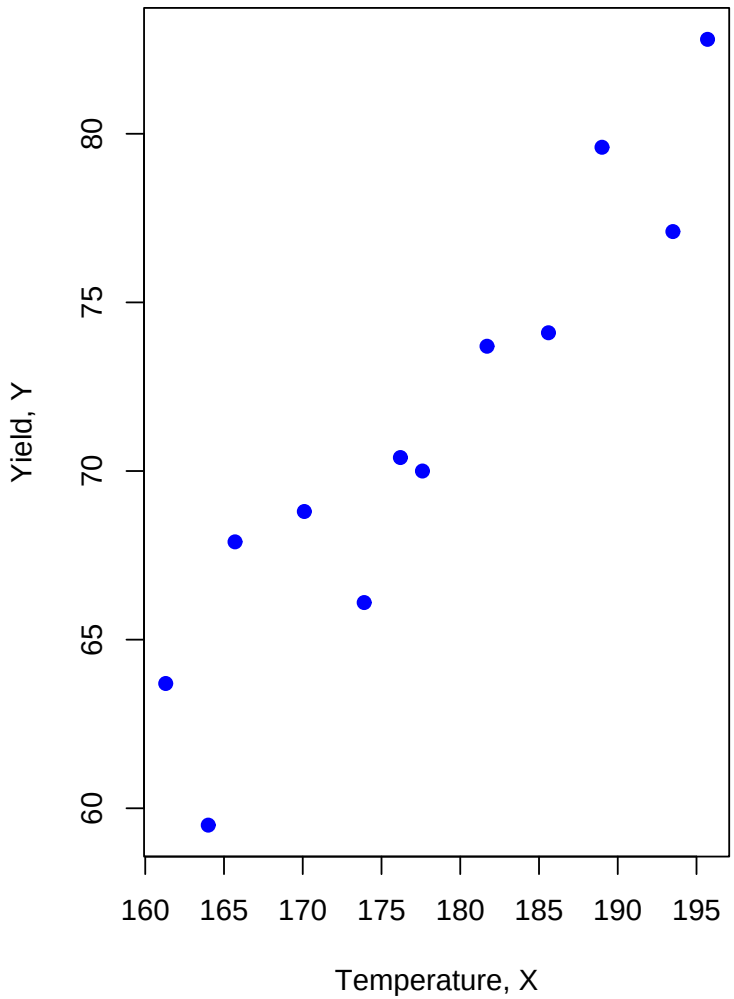
X	Y	Z	X	Y	Z
161.30	63.70	20.30	177.60	70.00	18.20
164.00	59.50	24.20	181.70	73.70	15.40
165.70	67.90	18.00	185.60	74.10	17.80
170.10	68.80	20.50	189.00	79.60	13.30
173.90	66.10	20.10	193.50	77.10	16.70
176.20	70.40	17.50	195.70	82.80	14.80

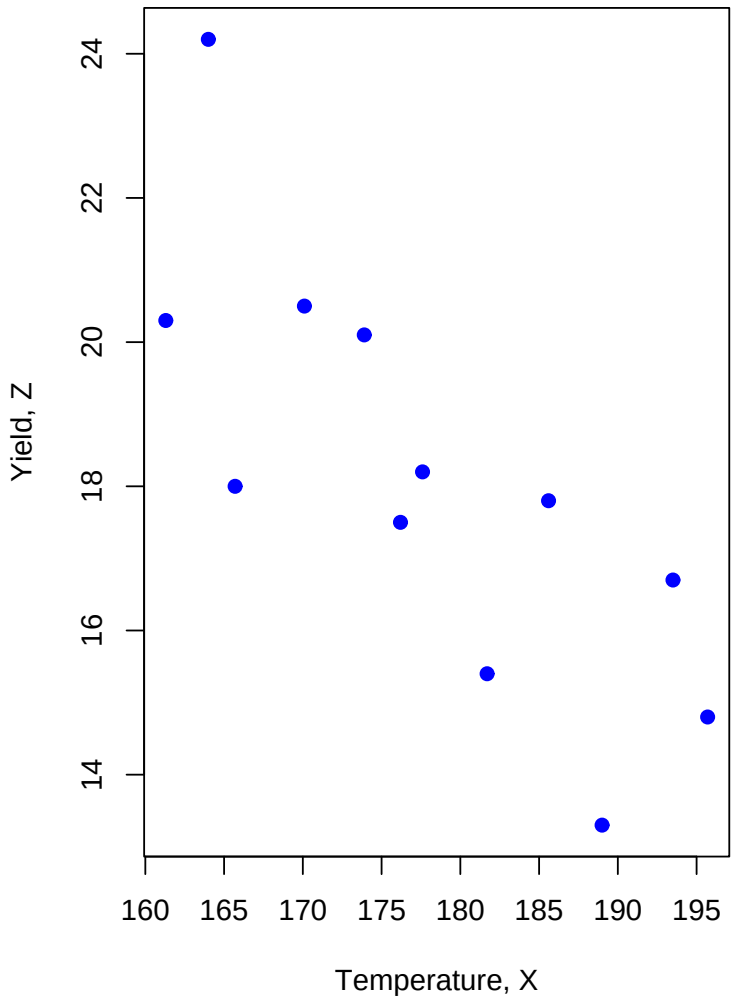
- Model:

$$Y_i = a + bx_i + e_i$$
$$Z_i = c + dx_i + f_i, \quad i = 1, \ldots, 12.$$

- a, b, c, d explain the relationships between the responses and the stimuli; $\{e_i, f_i\}$ are error quantities.

Yields of two products Y, Z as Temperature X varies





Prior judgements

- Separate runs 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.
- $e_i \sim \textbf{iid}(0, \sigma_e^2); f_i \sim \textbf{iid}(0, \sigma_f^2)$; $\mathrm{Cov}(e_i, f_i) = \sigma_{ef} \ \ \forall i$, $\mathrm{Cov}(e_i, f_j) = 0 \ \ \forall i \neq j$.
- 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 0.5.
- We specified the following expectations and covariances between the regression coefficients:

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

- Straightforward to construct a Bayes linear graphical model for this problem.

Bayes linear graphical model for the regression model (i, j)

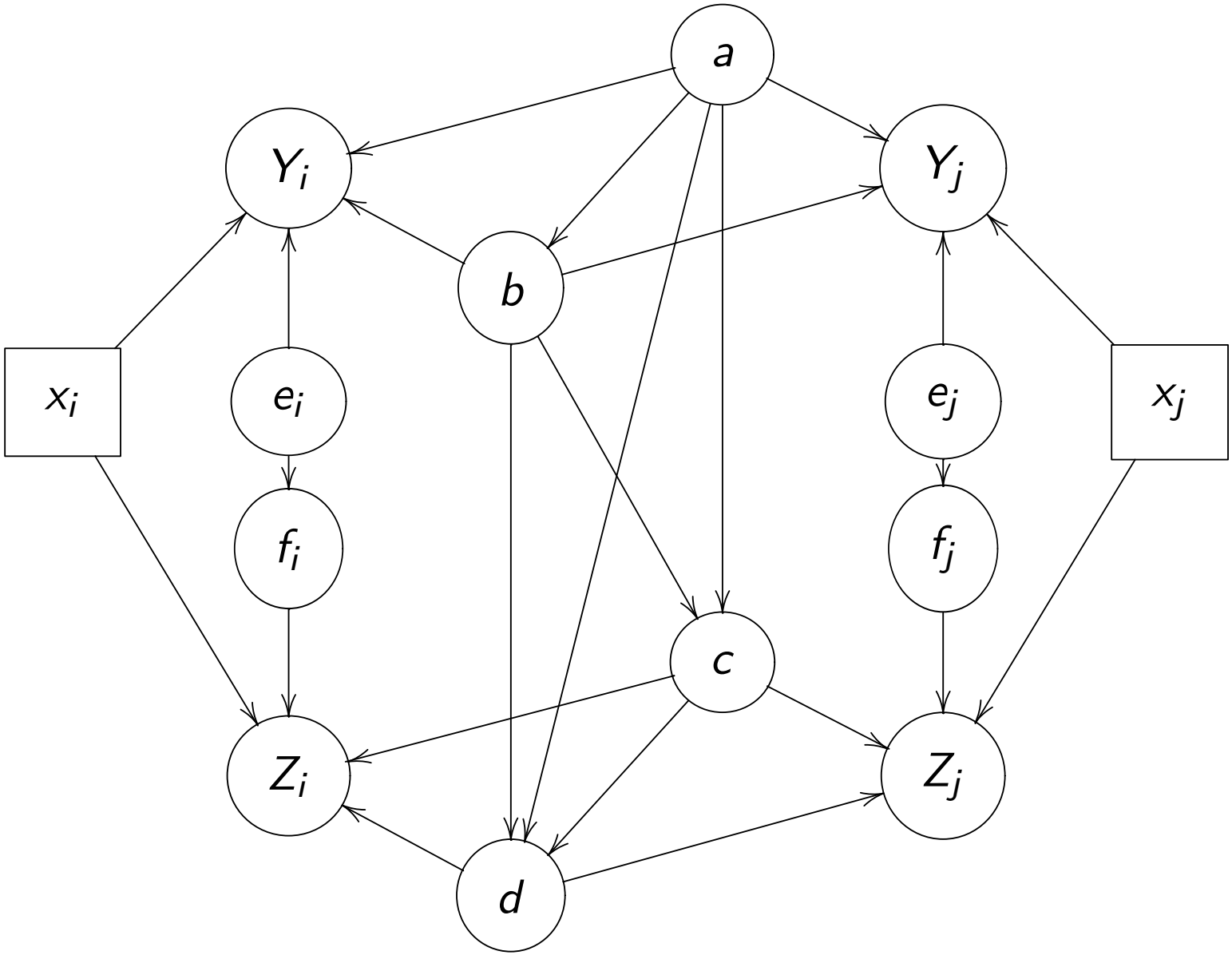
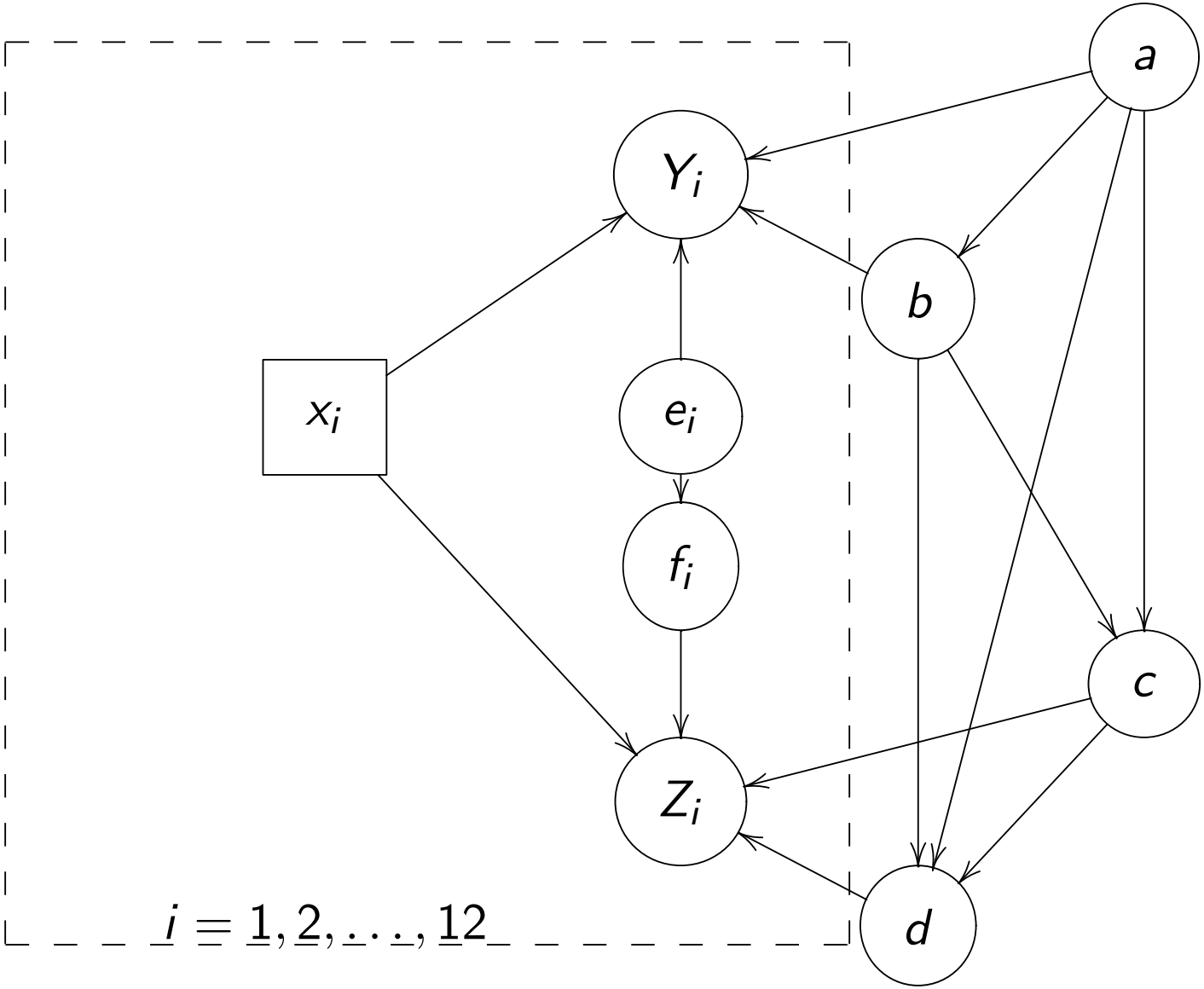


Plate representation



Operations on Bayes linear graphical models

- Fundamental notion is of (Bayes linear) separation,

$$\lfloor A \perp\!\!\!\perp B \rfloor \,/\, C,$$

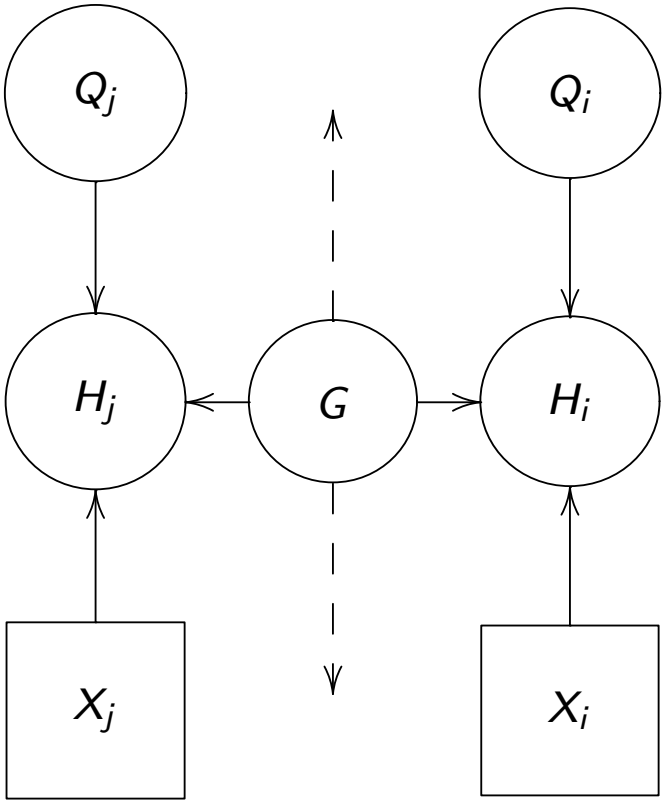
notation for collections A, B being separated by a collection C. Separation on the graph (DAG) is the property that for collections (nodes) A, B, C,

  - $\mathrm{E}_{C \cup A}(B) = \mathrm{E}_C(B)$;
  - $\mathrm{Var}_{C \cup A}(B) = \mathrm{Var}_C(B)$;
  - $\mathbb{T}_{B:C \cup A} = \mathbb{T}_{B:C}$.

That is, C is **Bayes linear sufficient** for A for adjusting B, i.e. if we want to revise beliefs for B knowing C and A, we can throw away A.

- $\lfloor A \perp\!\!\!\perp B \rfloor \,/\, C \iff \mathrm{Cov}(A, B) = \mathrm{Cov}(A, C)\mathrm{Var}(C)^{\dagger}\mathrm{Cov}(C, B)$.
- Belief separation is a generalized conditional independence property (Goldstein, 1990).
- BLGMs form the Bayes linear analogue of BBNs, with similar rules for node and arc operations, construction of junction trees and propagation of information (Wilkinson, 1998; Goldstein & Wilkinson, 2000; Goldstein & Wooff 2007).

Adjusting by a series of plates



- Revise beliefs over the collection of regression coefficients, $G = \{a, b, c, d\}$, by the pairs of measurements

$$H_i = \{Y_i, Z_i\}, \ i = 1, 2, \ldots, 12.$$

with errors $Q_i = \{E_i, F_i\}$ and design matrices X_i.

- That is, we may write the model as

$$H_i = X_i G + Q_i, \quad i = 1, 2, \ldots, n,$$

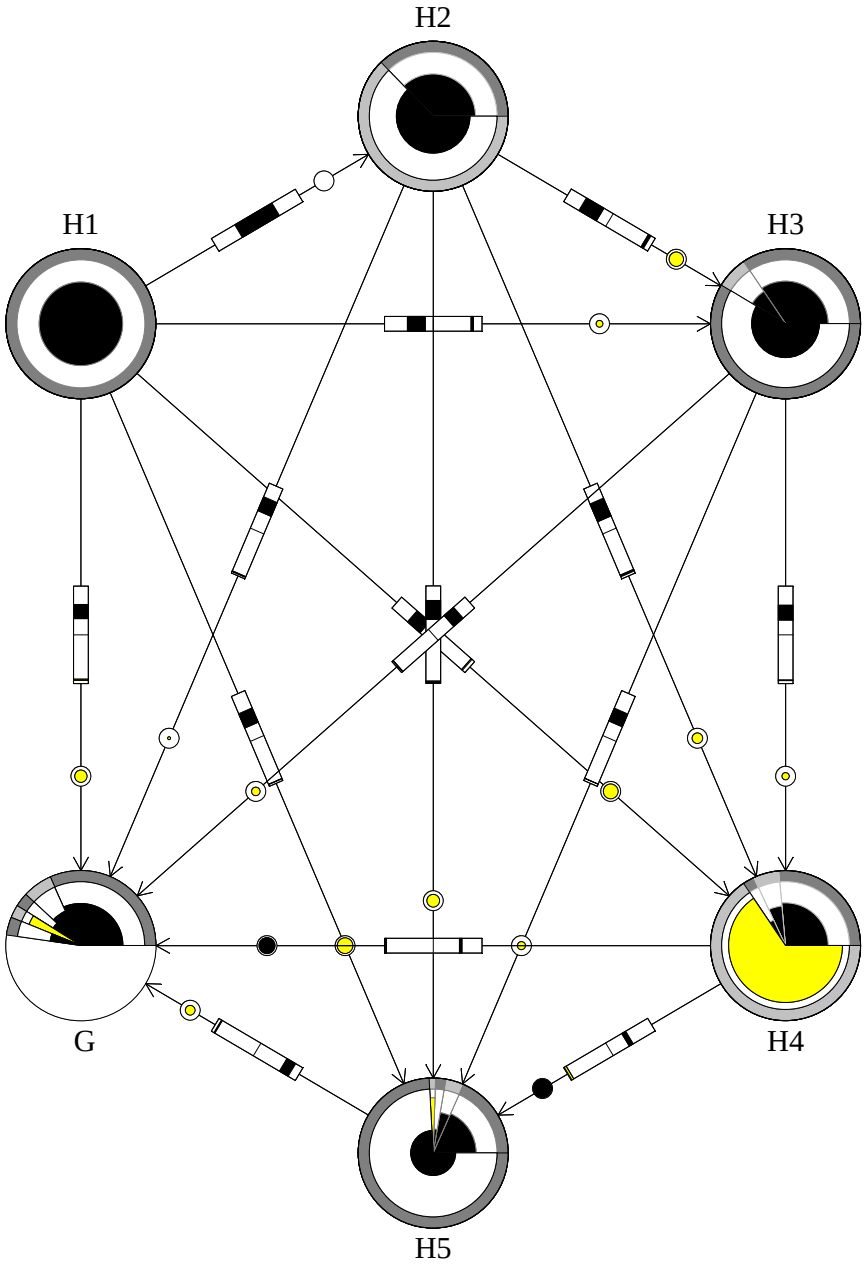
where $X_i = \begin{bmatrix} 1 & x_i & 0 & 0 \\ 0 & 0 & 1 & x_i \end{bmatrix}$.

General case

- Large class of linear template models

$$H_i = X_i G + Q_i, \quad i = 1, 2, \ldots, n,$$

  with $\mathrm{Var}(G) = \Sigma$ (pos. def. wlog) and $\mathrm{Var}(Q_i) = W_i$ (pos.def.)
- $\lfloor H_i \perp\!\!\!\perp H_j \rfloor \,/\, G$
- H_i is constructed from G, but belief revision of G by H_i requires arc reversals - this complicates the graph.



- Sequential adjustment of the parameter set G and subsequent observables $H_1, \ldots, H_5$
- Outer shading shows variance explained by successive sources
- Inner shading shows diagnostics:
  - No shading $\longrightarrow$ no surprising features
  - Black $\longrightarrow$ unexpectedly large changes in expectation
  - Yellow $\longrightarrow$ unexpectedly small changes in expectation
- Arc labels show information leaving/arriving plus implied diagnostics
- Circles on arcs show path correlations: whether information sources are consistent or contradictory
- Can use colour more effectively - this picture is intended to work in monochrome.

Questions

- For Goldstein & Wooff (2007), these belief revisions were constructed by brute force. We have $\lfloor H_i \perp\!\!\!\perp H_j \rfloor \,/\, G$, but not $\lfloor H_j \perp\!\!\!\perp G \rfloor \,/\, H_i$, which is what we would like to exploit.
- We had in mind then that there might be interesting symmetries to understand, for such linear templates, but ....
- .... our book was already 12 years late ☹
- So,
  - What symmetries, if any, may be exploited?
  - Can se say anything about sample size?
  - Can we say anything about designing the next experiment?

A sufficient set of observables

- Goldstein & Wooff (2007) show that there exists a minimal linear sufficient collection of quantities, termed the **heart of the transform** $\mathbb{H}(D/B)$, which carries all the information required to adjust B by D (and vice-versa). Similar result in classical multivariate statistics, e.g. seen in David Cox's talk this morning.
- The size of this collection is at most $\mathbf{rk}\{\mathrm{Var}(B)\}$. Such a collection may be determined by the eigenstructure of the projection from $[B]$ into $[D]$ and back again.
- For the adjustment of templates, this collection has a nice representation.
- Construct the vector observables

$$D_i = X_i^T W_i^{-1} H_i$$

and form their mean,

$$\bar{D} = \frac{1}{n} \sum_{i=1}^{n} D_i.$$

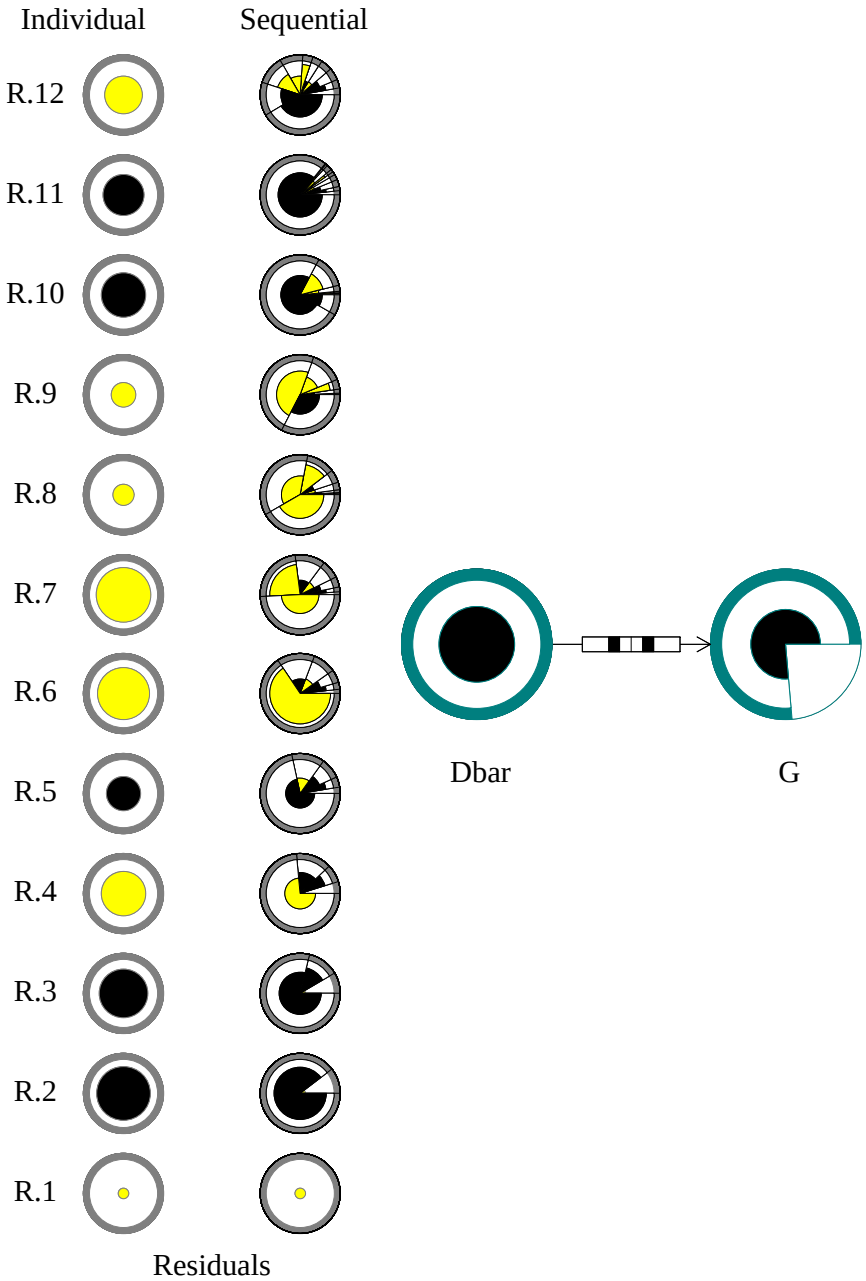
- These constructs allow a number of separations: $\lfloor H_i \perp\!\!\!\perp G \rfloor \,/\, D_i$, $\lfloor D_i \perp\!\!\!\perp G \rfloor \,/\, \bar{D}$, $\lfloor H \perp\!\!\!\perp G \rfloor \,/\, \bar{D}$. That is, $\bar{D}$ is Bayes linear sufficient for the observables for adjusting G.
- This is analogous (I think) to standard results for Gaussian models, classical and Bayesian.

Residuals

- We may construct alternative sets of residuals, for example

$$H_i - \mathrm{E}_{\bar{D}}(H_i), \ \ \forall i, \text{ and } D_i - \mathrm{E}_{\bar{D}}(D_i), \ \ \forall i.$$

- These are **ancillary** statistics: they can tell us nothing about the parameter collection G, but they might allow us to diagnose problems with our prior formulation.
- It is natural to examine diagnostics on the observed residuals. For problems with a natural time-ordering, it may also be useful to inspect the sequential adjustments of these residuals by themselves.
- Several natural ways of plotting these residual collections, for example with node areas proportional to prior variation.
- We do not normally show arcs between the residual collections, to avoid clutter.
- The residuals may be used for variance learning and thence to two-stage Bayes linear analysis.



- Left-hand residuals show observed values of $H_i - \mathrm{E}_{\bar{D}}(H_i)$. 12 2-dimensional residuals in a 20-dimensional space.
- Right-hand residuals show the sequential adjustment of the residual quantities by themselves.
- $\bar{D}$ is 4-dimensional, observed, and used to update parameter set G.
- G has 76.4% of its uncertainty explained.
- The data imply quite large changes in expectation relative to variation explained.
- Residuals should show random diagnostic patterns - do they? Arguably some edge effects.
- Unexpectedly large changes in expectation at the edges, unexpectedly small changes in the centre. Might suggest that assumptions about the error quantities are inappropriate.

Digression: second-order exchangeable sequences

- Consider a second-order exchangeable vector sequence of observable random quantities $U_1, U_2, \ldots$, with $\mathrm{E}(U_i) = \mu$, $\mathrm{Var}(U_i) = \Phi$, $\mathrm{Cov}(U_i, U_j) = \Psi$.
- Consider a further r-dimensional vector V, and that we wish to revise beliefs over V using the n observables $U_1, U_2, \ldots, U_n$.
- Let $\bar{U} = \frac{1}{n}\sum_{i=1}^{n} U_i$, and let U^n be the collection $\{U_1, U_2, \ldots, U_n\}$
- Then $\lfloor V \perp\!\!\!\perp U^n \rfloor \,/\, \bar{U}$, i.e. the sample mean is Bayes linear sufficient for the both the full collection U^n and the individual quantities U_i for adjusting V.
- Construct the resolution transform $\mathbb{T}_{U_1:V}$ and determine its eigenvalues $\lambda_1, \ldots$ and corresponding eigenvectors. Construct the diagonal matrix of eigenvalues

$$\Lambda_n = \mathbf{diag}\{\frac{n\lambda_1}{1+(n-1)\lambda_1}, \ldots, \frac{n\lambda_r}{1+(n-1)\lambda_r}\},$$

- Then $\mathbb{T}_{V:U^n} = \mathrm{Var}(V)^{\dagger}\mathrm{Cov}(V, \bar{U})\Lambda_n\mathrm{Cov}(\bar{U}, V)$.
- Implication: everything about the belief revision for V can be determined from the average $\bar{U}$ and the sample size n, so this is the essence of how we may exploit the symmetry in beliefs over the observables.
- The residual quantities $U_i - \mathrm{E}_{\bar{U}}(U_i)$ are useless for revising beliefs over V, but can and should be explored for anomalies.

Design and calibration

- Consider the hypothetical second-order exchangeable series of vectors $A_1, A_2, \ldots$, defined as $A_i = G + R_i$, with prior beliefs $\mathrm{E}(G)$, $\mathrm{E}(R_i) = 0$, $\mathrm{Var}(G) = \Sigma$,

$$\mathrm{Var}(R_i) = (\frac{1}{n}\Omega)^{\dagger} = (\frac{1}{n}\sum_{i=1}^{n} X_i^T W_i^{-1} X_i)^{\dagger} \; \forall i,$$

  and other correlations being zero.
- It can be established that

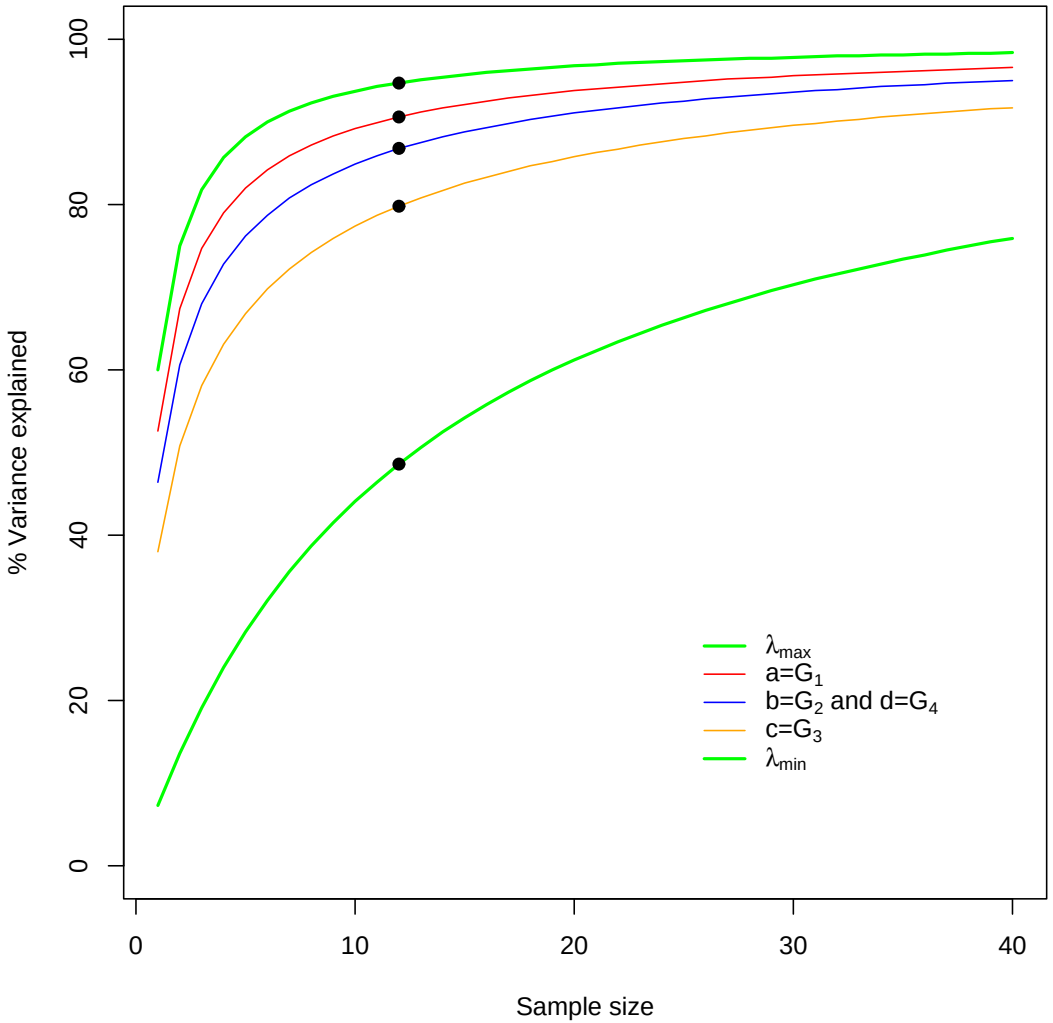
$$\mathrm{E}_H(G) = \mathrm{E}_A(G), \;\; \mathrm{Var}_H(G) = \mathrm{Var}_A(G), \;\; \mathbb{T}_{G:H} = \mathbb{T}_{G:A}.$$

- That is, the sequence of observables $H_1, H_2, \ldots, H_n$ and its implications for learning about G is consistent with the existence of an infinite exchangeable sequence $A_1, A_2, \ldots$ with the given belief specifications.
- Therefore, we may use the resolution transform $\mathbb{T}_{G:A}$ to explore sample size considerations, on the assumption that the mean weighted precision matrix $\frac{1}{n}\sum_{i=1}^{n} X_i^T W_i^{-1} X_i$ is "typical" of the weighted precision matrices

$$X_{n+1}^T W_{n+1}^{-1} X_{n+1}, \;\; X_{n+2}^T W_{n+2}^{-1} X_{n+2} \ldots$$

  at additional design points.

Effect of increasing sample size on variance explained



- Solve the generalized eigenvalue problem

$$\Sigma V = (\Sigma + n[\sum_{i=1}^{n} X_i^T W_i^{-1} X_i]^{\dagger}) V \Lambda_1,$$

to obtain eigenvalues Λ_1 for a notional sample size $m = 1$ from the sequence $A_1, A_2, \ldots$

- $\Lambda_m = m\Lambda_1[(m-1)\Lambda_1 + I]^{-1}$, we only need to compute the $m = 1$ structure, everything else can be deduced.
- Variance resolutions for original parameters are bounded by largest and smallest eigenvalues for each sample size.

Hierarchical parameter set

- Suppose that the regression model is more complicated, namely that our model is

$$H_i = X_i G_i + Q_i$$

where G_i is a second-order exchangeable sequence of parameter vectors such that

$$G_i = \mathcal{M}(G) + \mathcal{R}_i(G),$$

with $\mathrm{E}(\mathcal{M}(G)) = \mathrm{E}(G)$, $\mathrm{Var}(\mathcal{R}_i(G)) = \Gamma - \Sigma$, $\mathrm{Var}(\mathcal{M}(G)) = \Sigma$, with Γ and $\Gamma - \Sigma$ pos. def. We may then rewrite our model as

$$H_i = X_i \mathcal{M}(G) + Q_i^*,$$

where $Q_i^* = Q_i + \mathcal{R}_i(G)$ and $\mathrm{Var}(Q_i^*) = W_i + (\Gamma - \Sigma) = W_i^*$. This is functionally identical to the previous model, and so can be treated in the same way.

Design of next observation

- Suppose that we may afford a further observation, H_{n+1}.
- The variance explained by the extra observation is

$$\mathrm{Var}_{\bar{D}}(G) - \mathrm{Var}_{\bar{D}\cup H_{n+1}}(G) = \Sigma\Omega_n[(\Sigma^{-1}+\Omega_n)^{-1} - (\Sigma^{-1}+\Omega_n+X_{n+1}^T W_{n+1}^{-1} X_{n+1})^{-1}]\Omega_n\Sigma,$$

  with $\Sigma = \mathrm{Var}(G)$ and $\Omega_n = \sum_{j=1}^{n} X_j^T W_j^{-1} X_j$.
- Pleasing quadratic form, but most useful only when W_{n+1} does not depend on X_{n+1}, otherwise gets tricky.
- Work to do ....

Conclusions

- Some re-inventing of the wheel - but using linear tools!
- Leads to a clear(er) understanding of the adjustment through underlying separations
- Lots of interesting (to me) issues still to explore
  - Variance learning
  - Teasing out of the null part of the residual space
  - General representations of linear transformations and decompositions over graphs
- M. Goldstein & D.A. Wooff (2007) Bayes linear statistics: theory and methods. Chichester: Wiley.