Generalised Bayesian graphical modelling utilising Bayes linear kinematics

Simon Shaw

Department of Mathematical Sciences
University of Bath

joint work with

Michael Goldstein
University of Durham

9 July 2008

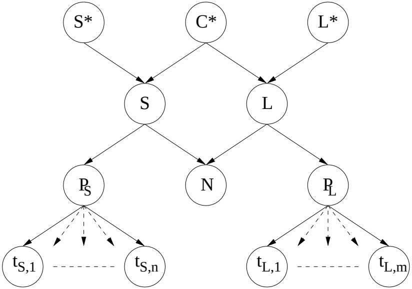


Figure: A Bayesian graphical model modelling a software action. S are short numbers, L long numbers. Note that $(P_S \perp\!\!\!\perp P_L)|S, L$.

Modelling complex systems

- decision that $(P_S \perp\!\!\!\perp P_L) | S, L$ is often a modelling simplification and may not be merited in practice
- want to introduce a richer dependency structure into the model
- consider that full joint specification of P_S and P_L is beyond our full capability or desirability
- seek to utilise the strength of the Bayes linear methodology to introduce a direct dependency between these variables in the form of a partial specification whilst maintaining the probabilistic structure of the model
- result is a generalised Bayesian graphical model featuring different levels of specification in different areas of the model

Modelling complex systems

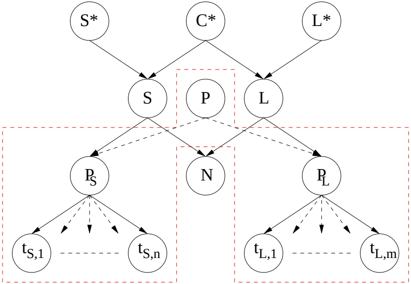


Figure: A Bayesian graphical model modelling a software action. S are short numbers, L long numbers. Note that we no longer have $(P_S \perp\!\!\!\perp P_L) | S, L$.

Modelling complex systems

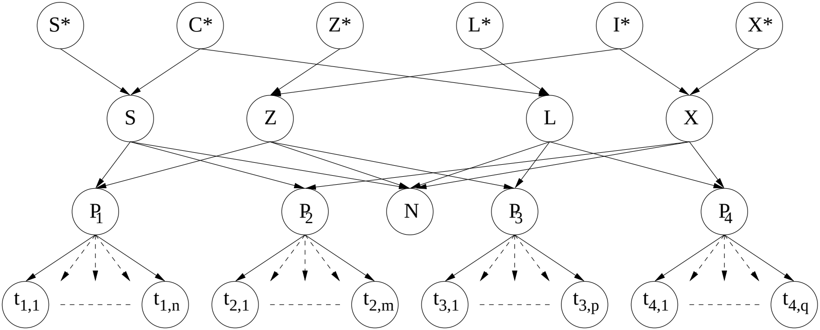


Figure: Extended Bayesian graphical model modelling a software action. S are short numbers, L are long numbers, Z are numbers commencing with a zero and X are numbers not commencing with a zero.

Suppose that $\mathcal{X}$, $\mathcal{Y}$ are two collections of random quantities of interest

- view $\mathcal{X}$ as a $r \times 1$ vector and specify the prior mean vector and prior variance matrix and collect these together as

$$\lfloor \mathcal{X} \rceil \quad = \quad \{E(\mathcal{X}), Var(\mathcal{X})\}$$

- considering $\mathcal{Y}$ as a $s \times 1$ vector, specifying $E(\mathcal{Y})$, $Var(\mathcal{Y})$ yields the specification $\lfloor \mathcal{Y} \rceil$

  specification of the covariance matrix $Cov(\mathcal{X}, \mathcal{Y})$ allows the construction of $\lfloor \mathcal{B} \rceil$, where $\mathcal{B} = \mathcal{X} \cup \mathcal{Y}$.

Bayes linear analysis

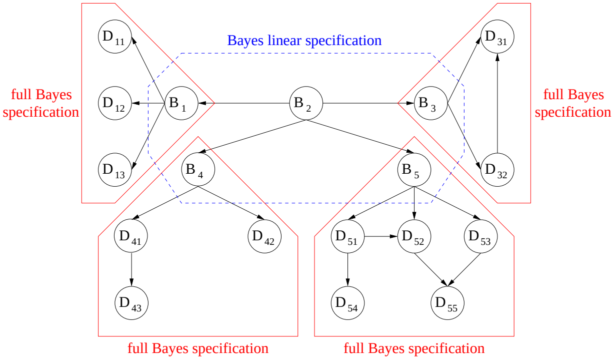
Adjusted mean and variance

The Bayes linear adjusted mean and variance for $\mathcal{Y}$ given $\mathcal{X}$ are given by

$$\begin{aligned} E_{\mathcal{X}}(\mathcal{Y}) &= E(\mathcal{Y}) + W_{\mathcal{X}(\mathcal{Y})}\{\mathcal{X} - E(\mathcal{X})\}; \\ Var_{\mathcal{X}}(\mathcal{Y}) &= Var(\mathcal{Y}) - Cov(\mathcal{Y}, \mathcal{X})W^T_{\mathcal{X}(\mathcal{Y})}, \end{aligned}$$

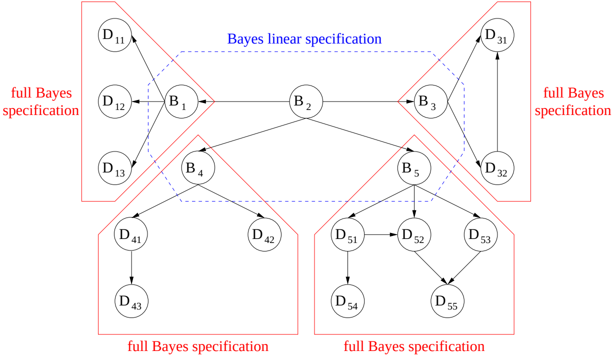
where $W_{\mathcal{X}(\mathcal{Y})} = Cov(\mathcal{Y}, \mathcal{X})Var^{-1}(\mathcal{X})$.

Bayes linear Bayes (BLB) graphical model



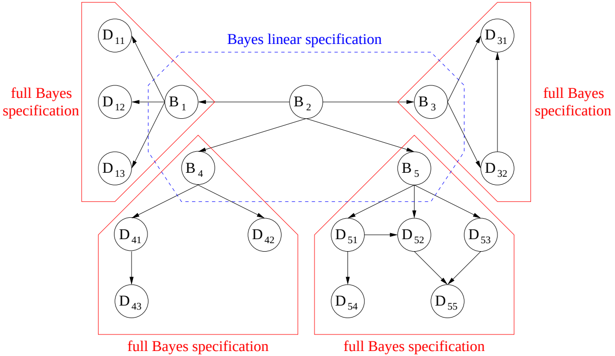
- Define the Bayes linear Bayes (BLB) graphical model, a mixture of Bayesian and Bayes linear graphical models, as follows:

Bayes linear Bayes (BLB) graphical model



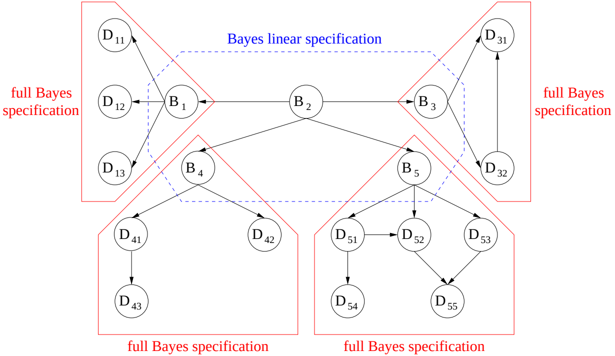
[B1] The graph $\mathcal{G}_{\mathcal{B}} = (\mathcal{B}, E_{\mathcal{B}})$, where $\mathcal{B} = \{\mathcal{B}_1, \ldots, \mathcal{B}_n\}$, is a Bayes linear graphical model. Each node $\mathcal{B}_i$ is a random vector. We make a full second order prior specification over all elements of $\mathcal{B}$, denoted $\lfloor \mathcal{B} \rceil_0$.

Bayes linear Bayes (BLB) graphical model



[B2] Certain nodes $\mathcal{B}_g$ are also elements of disjoint Bayesian graphical models $\mathcal{G}_{V_g} = (V_g, E_{V_g})$, where $V_g = \mathcal{B}_g \cup \mathcal{D}_g$, subject to the conditions that for any $\mathcal{D}_{gi} \in \mathcal{D}_g$, $pa(\mathcal{D}_{gi}) \subset V_g$ and $ch(\mathcal{D}_{gi}) \subset \mathcal{D}_g$. We make a full joint probabilistic specification over the elements of each $\mathcal{G}_{V_g}$ denoted $f_0(v_g)$.

Bayes linear Bayes (BLB) graphical model



[B3] The belief specification is completed by the condition, for each g, that, given $\mathcal{B}_g$, the collection $\mathcal{D}_g$ is conditionally independent of all the remaining elements of $\mathcal{B}$ and of all elements of each $\mathcal{D}_h, h \neq g$.

Bayes linear kinematics

- Bayes linear kinematics is the Bayes linear analogue of Richard Jeffrey's probability kinematics
- Let $\mathcal{B} = \mathcal{X} \cup \mathcal{Y}$ be a collection of random quantities with a second-order specification, $\lfloor \mathcal{B} \rceil_o = \{E_o(\mathcal{B}),\ Var_o(\mathcal{B})\}$, attached.
- rather than observing $\mathcal{X}$, we instead only receive partial information which changes our beliefs about $\mathcal{X}$ in some generalised way, without affecting our adjusted beliefs for $\mathcal{Y}$ given $\mathcal{X}$
  - i.e. were we now to learn the value of $\mathcal{X}$, then the preceding information would be deemed irrelevant to the subsequent belief adjustment
- new specification, $\lfloor \mathcal{X} \rceil_n = \{E_n(\mathcal{X}),\ Var_n(\mathcal{X})\}$, for $\mathcal{X}$ is given

Bayes linear kinematics

Bayes linear kinematics

We say the new specification $\lfloor\mathcal{B}\rceil_n = \{E_n(\mathcal{B}),\ Var_n(\mathcal{B})\}$ comes from the old specification $\lfloor\mathcal{B}\rceil_o$ by Bayes linear kinematics (BLK) on $\lfloor\mathcal{X}\rceil_n$ if the following two Bayes linear sufficiency conditions are satisfied

$$\begin{aligned} E_{o\mathcal{X}}(\mathcal{Y}) &= E_{n\mathcal{X}}(\mathcal{Y}); \\ Var_{o\mathcal{X}}(\mathcal{Y}) &= Var_{n\mathcal{X}}(\mathcal{Y}). \end{aligned}$$

- $E_{o\mathcal{X}}(\mathcal{Y})$, $E_{n\mathcal{X}}(\mathcal{Y})$ denote the old and new assignments for the adjusted expectation, evaluated using the old and new specifications $\lfloor\mathcal{B}\rceil_o$, $\lfloor\mathcal{B}\rceil_n$ respectively.
- $Var_{o\mathcal{X}}(\mathcal{Y})$, $Var_{n\mathcal{X}}(\mathcal{Y})$ are the old and new adjusted variance

Bayes linear kinematics

Theorem

The Bayes linear sufficiency conditions $E_{o\mathcal{X}}(\mathcal{Y}) = E_{n\mathcal{X}}(\mathcal{Y})$; $Var_{o\mathcal{X}}(\mathcal{Y}) = Var_{n\mathcal{X}}(\mathcal{Y})$ *are equivalent to the requirement that*

$$\begin{aligned} E_n(\mathcal{Y}) &= E_o(\mathcal{Y}) + W_{o,\mathcal{X}(\mathcal{Y})}\{E_n(\mathcal{X}) - E_o(\mathcal{X})\}; \\ Var_n(\mathcal{Y}) &= Var_{o\mathcal{X}}(\mathcal{Y}) + W_{o,\mathcal{X}(\mathcal{Y})}Var_n(\mathcal{X})W^T_{o,\mathcal{X}(\mathcal{Y})}; \\ Cov_n(\mathcal{Y},\mathcal{X}) &= W_{o,\mathcal{X}(\mathcal{Y})}Var_n(\mathcal{X}), \end{aligned}$$

where $W_{o,\mathcal{X}(\mathcal{Y})} = Cov_o(\mathcal{Y},\mathcal{X})Var_o^{-1}(\mathcal{X})$.

1. if $\mathcal{X}$ is observed, so that $E_n(\mathcal{X}) = \mathcal{X}$ and $Var_n(\mathcal{X}) = 0$, then $E_n(\mathcal{Y}) = E_{o\mathcal{X}}(\mathcal{Y})$ and $Var_n(\mathcal{Y}) = Var_{o\mathcal{X}}(\mathcal{Y})$
   - Bayes linear kinematics provides a generalisation of Bayes linear adjustment to cover the case where the change in belief over $\mathcal{X}$ may not result in $\mathcal{X}$ being known with certainty.

Bayes linear kinematics

2. if $\mathcal{D}$ is a further vector of random quantities for which the partial correlations between $\mathcal{Y}$ and $\mathcal{D}$ given $\mathcal{X}$ are zero, so that

$$Cov_o(\mathcal{Y} - E_{o\mathcal{X}}(\mathcal{Y}), \mathcal{D} - E_{o\mathcal{X}}(\mathcal{D})) \;\; = \;\; 0.$$

   we say that $\mathcal{Y}$ is separated from $\mathcal{D}$ by $\mathcal{X}$ in $\lfloor \mathcal{C} \rceil_o$, written $(\mathcal{Y} \perp\!\!\!\perp_o \mathcal{D})|\mathcal{X}$, where $\mathcal{C} = \mathcal{B} \cup \mathcal{D}$. In this case, if we obtain $\lfloor \mathcal{X} \rceil_n$ by Bayes linear adjustment of $\mathcal{X}$ by $\mathcal{D}$, then the Bayes linear sufficiency conditions hold automatically

3. Not all $\lfloor \mathcal{B} \rceil_n$ can be obtained from $\lfloor \mathcal{B} \rceil_o$ by linear fitting on some $\mathcal{D}$: linear fitting shrinks the variances but there is no constraint that $Var_n(\mathcal{X}) \leq Var_o(\mathcal{X})$

Bayes linear kinematics: consistency check

- suppose that $\mathcal{A}$ is a further vector of random quantities
- let $\mathcal{B}^* = \mathcal{A} \cup \mathcal{X} \cup \mathcal{Y}$, with $\lfloor \mathcal{B}^* \rceil_o$, $\lfloor \mathcal{B}^* \rceil_n$ respectively denoting the old and new second-order specification for $\mathcal{B}^*$.

Theorem

If $\lfloor \mathcal{B}^* \rceil_n$ is obtained from $\lfloor \mathcal{B}^* \rceil_o$ by BLK on $\lfloor \mathcal{X} \rceil_n$ then $\lfloor \mathcal{B}^* \rceil_n$ is equivalently obtained from $\lfloor \mathcal{B}^* \rceil_o$ by BLK on $\lfloor \mathcal{B} \rceil_n$, where $\mathcal{B} = \mathcal{X} \cup \mathcal{Y}$.

- the kinematic for B^* may be constructed sequentially by first using $\lfloor \mathcal{X} \rceil_n$ to construct $\lfloor \mathcal{B} \rceil_n$ and then using $\lfloor \mathcal{B} \rceil_n$ to obtain $\lfloor \mathcal{B}^* \rceil_n$

Bayes linear kinematics: successive revisions

Theorem

If $\lfloor \mathcal{B}^* \rceil_1$ is obtained from $\lfloor \mathcal{B}^* \rceil_0$ by BLK on $\lfloor \mathcal{X} \rceil_1$ and $\lfloor \mathcal{B}^* \rceil_{12}$ is obtained from $\lfloor \mathcal{B}^* \rceil_1$ by BLK on $\lfloor \mathcal{Y} \rceil_{12}$ then $\lfloor \mathcal{B}^* \rceil_{12}$ is equivalently obtained from $\lfloor \mathcal{B}^* \rceil_0$ by BLK on $\lfloor \mathcal{B} \rceil_{12}$.

$$\begin{tikzcd}
\lfloor \mathcal{B}^* \rceil_0 \arrow[r, "\lfloor \mathcal{X} \rceil_1"] \arrow[dr, "\lfloor \mathcal{B} \rceil_{12}"'] & \lfloor \mathcal{B}^* \rceil_1 \arrow[d, "\lfloor \mathcal{Y} \rceil_{12}"] \\
 & \lfloor \mathcal{B}^* \rceil_{12}
\end{tikzcd}$$

Bayes linear kinematics: commutativity

$$
\begin{tikzcd}
\lfloor\mathcal{B}\rceil_0 \arrow[r, "\lfloor\mathcal{X}\rceil_1"] \arrow[dd, "\lfloor\mathcal{Y}\rceil_2"'] & \lfloor\mathcal{B}\rceil_1 \arrow[d, "\lfloor\mathcal{Y}\rceil_{12}"] \\
 & \lfloor\mathcal{B}\rceil_{12} \\
\lfloor\mathcal{B}\rceil_2 \arrow[r, "\lfloor\mathcal{X}\rceil_{21}"'] & \lfloor\mathcal{B}\rceil_{21}
\end{tikzcd}
$$

Key question: when does $\lfloor\mathcal{B}\rceil_{12} = \lfloor\mathcal{B}\rceil_{21}$?

We now consider successive belief revisions using Bayes linear kinematics and explore the constraints imposed upon our specifications if the order of the revisions may be reversed

Bayes linear kinematics: commutativity

Theorem

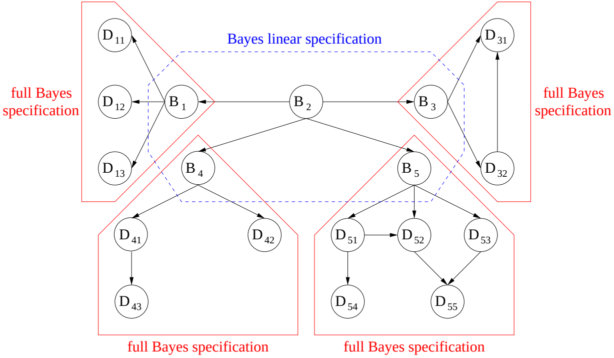
There is a unique solution to $\lfloor\mathcal{B}^*\rceil_{12} = \lfloor\mathcal{B}^*\rceil_{21}$ *for all* $B^* \in \langle\mathcal{B}^*\rangle$ *if and only if* $Var_1^{-1}(\mathcal{X})+ Var_2^{-1}(\mathcal{X})- Var_0^{-1}(\mathcal{X})$ *is positive definite if and only if* $Var_1^{-1}(\mathcal{Y}) + Var_2^{-1}(\mathcal{Y})- Var_0^{-1}(\mathcal{Y})$ *is positive definite. The unique solution may be expressed as*

$$\begin{aligned} Var_{12}(\mathcal{W}) &= \{Var_1^{-1}(\mathcal{W}) + Var_2^{-1}(\mathcal{W}) - Var_0^{-1}(\mathcal{W})\}^{-1}; \\ E_{12}(\mathcal{W}) &= \{Var_1^{-1}(\mathcal{W}) + Var_2^{-1}(\mathcal{W}) - Var_0^{-1}(\mathcal{W})\}^{-1} \\ &\quad \{Var_1^{-1}(\mathcal{W})E_1(\mathcal{W}) + Var_2^{-1}(\mathcal{W})E_2(\mathcal{W}) - \\ &\quad Var_0^{-1}(\mathcal{W})E_0(\mathcal{W})\}; \end{aligned}$$

where $\mathcal{W}$ *is equal to either* $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{B}$ *or* $\mathcal{B}^*$.

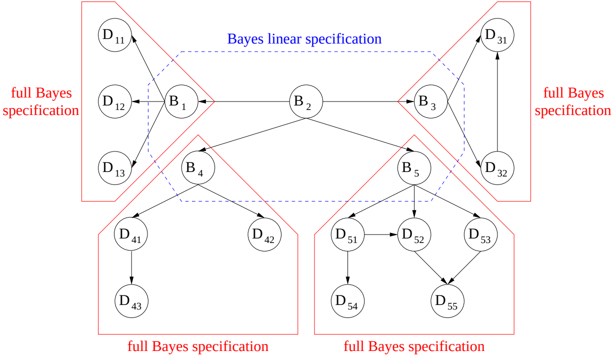
- under commutativity, $\lfloor\mathcal{W}\rceil_0$, $\lfloor\mathcal{W}\rceil_1$, $\lfloor\mathcal{W}\rceil_2$ are sufficient for calculating $\lfloor\mathcal{W}\rceil_{12}$.

Bayes linear Bayes (BLB) graphical model



- Observe some data in $\mathcal{D}_g$ and update beliefs about $\mathcal{B}_g$ using full conditioning
- Hence, obtain a new mean and variance for $\mathcal{B}_g$

Bayes linear Bayes (BLB) graphical model



- propagate this new specification over $\mathcal{B}$ using Bayes linear kinematics
- use successive Bayes linear kinematics to update over $\mathcal{B}$ from observations in multiple $\mathcal{D}_g$s

Bayes linear Bayes (BLB) graphical model

- suppose that we observe $\mathcal{D}^*_g \subseteq \mathcal{D}_g$
- construct $\lfloor \mathcal{B}_g \rceil_g = \{E_g(\mathcal{B}_g),\ Var_g(\mathcal{B}_g)\}$ where

$$\begin{aligned} E_g(\mathcal{B}_g) &= E_0(\mathcal{B}_g | \mathcal{D}^*_g); \\ Var_g(\mathcal{B}_g) &= Var_0(\mathcal{B}_g | \mathcal{D}^*_g) \end{aligned}$$

- impose the Bayes linear sufficiency conditions $E_{0\mathcal{B}_g}(\mathcal{B}) = E_{g\mathcal{B}_g}(\mathcal{B})$ and $Var_{0\mathcal{B}_g}(\mathcal{B}) = Var_{g\mathcal{B}_g}(\mathcal{B})$
- hence,

$$\begin{aligned} E_g(\mathcal{B}) &= E_0(\mathcal{B}) + W_{0,\mathcal{B}_g(\mathcal{B})}\{E_0(\mathcal{B}_g|\mathcal{D}^*_g) - E_0(\mathcal{B}_g)\}; \\ Var_g(\mathcal{B}) &= Var_{0\mathcal{B}_g}(\mathcal{B}) + W_{0,\mathcal{B}_g(\mathcal{B})} Var_0(\mathcal{B}_g|\mathcal{D}^*_g) W^T_{0,\mathcal{B}_g(\mathcal{B})}, \end{aligned}$$

  with $W_{0,\mathcal{B}_g(\mathcal{B})} = Cov_0(\mathcal{B}, \mathcal{B}_g) Var_0^{-1}(\mathcal{B}_g)$
- Bayes linear kinematics therefore provides a methodology for embedding a full probability update into a Bayes linear model

Bayes linear Bayes (BLB) graphical model

- use a series of sequential commutative Bayes linear kinematic updates to revise our beliefs over $\mathcal{B} = \{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_n\}$ following observation in multiple sources

$$
\begin{tikzcd}
\lfloor \mathcal{B} \rceil_0 \arrow[r, "\lfloor \mathcal{B}_1 \rceil_1"] \arrow[d, "\lfloor \mathcal{B}_2 \rceil_2"'] \arrow[dr, "\lfloor \mathcal{B}_1 \cup \mathcal{B}_2 \rceil_{12}"] & \lfloor \mathcal{B} \rceil_1 \arrow[d, "\lfloor \mathcal{B}_2 \rceil_{12}"] \\
\lfloor \mathcal{B} \rceil_2 \arrow[r, "\lfloor \mathcal{B}_1 \rceil_{21}"'] & \lfloor \mathcal{B} \rceil_{12}
\end{tikzcd}
$$

Bayes linear Bayes (BLB) graphical model

- use a series of sequential commutative Bayes linear kinematic updates to revise our beliefs over $\mathcal{B} = \{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_n\}$ following observation in multiple sources

$$
\begin{tikzcd}
\lfloor \mathcal{B} \rceil_0 \arrow[r, "\lfloor \mathcal{B}_1 \cup \mathcal{B}_2 \rceil_{12}"] \arrow[d, "\lfloor \mathcal{B}_3 \rceil_3"'] \arrow[dr, "\lfloor \mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3 \rceil_{123}"'] & \lfloor \mathcal{B} \rceil_{12} \arrow[d, "\lfloor \mathcal{B}_3 \rceil_{123}"] \\
\lfloor \mathcal{B} \rceil_3 \arrow[r, "\lfloor \mathcal{B}_1 \cup \mathcal{B}_2 \rceil_{312}"'] & \lfloor \mathcal{B} \rceil_{123}
\end{tikzcd}
$$

Bayes linear Bayes (BLB) graphical model

- consider revising our beliefs over $\mathcal{B}$ following observation of the collection $\mathcal{D}^* = \cup_{k=1}^{s} \mathcal{D}^*_{g_k}$ where $g_k \in \{1, \ldots, n\}$ and if $k \neq k'$, $g_k \neq g_{k'}$ and each $\mathcal{D}^*_{g_k} \subseteq \mathcal{D}_{g_k}$
- adding the data sequentially using commutative Bayes linear kinematics updates yields $\lfloor \mathcal{B} \rceil_{g[s]}$; $\lfloor \mathcal{B} \rceil_{g[s]}$ represents revised beliefs over $\mathcal{B}$ incorporating the data, $\mathcal{D}^*$

$$
\begin{aligned}
Var^{-1}_{g[s]}(\mathcal{B}) &= \sum_{k=1}^{s} Var^{-1}_{g_k}(\mathcal{B}) - (s-1) Var^{-1}_{0}(\mathcal{B}); \\
Var^{-1}_{g[s]}(\mathcal{B}) E_{g[s]}(\mathcal{B}) &= \sum_{k=1}^{s} Var^{-1}_{g_k}(\mathcal{B}) E_{g_k}(\mathcal{B}) - \\
&\qquad (s-1) Var^{-1}_{0}(\mathcal{B}) E_0(\mathcal{B}).
\end{aligned}
$$

Local computation

Local computation theorems

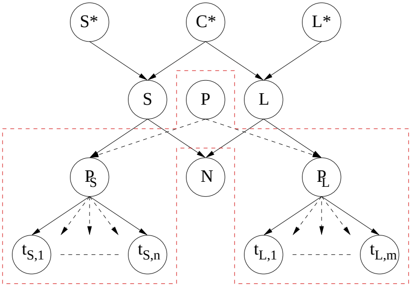
1. Let $\mathcal{B}^* = \mathcal{X} \cup \mathcal{Y} \cup \mathcal{A}$. If $\lfloor \mathcal{B}^* \rceil_1$ is obtained from $\lfloor \mathcal{B}^* \rceil_0$ by BLK on $\lfloor \mathcal{X} \rceil_1$ and $(\mathcal{A} \perp\!\!\!\perp_0 \mathcal{X}) \,|\, \mathcal{Y}$ then $\lfloor \mathcal{A} \cup \mathcal{Y} \rceil_1$ is obtained from $\lfloor \mathcal{A} \cup \mathcal{Y} \rceil_0$ by BLK on $\lfloor \mathcal{Y} \rceil_1$
2. Let $\mathcal{C} = \mathcal{F} \cup \mathcal{G} \cup \mathcal{H}$. If $\lfloor \mathcal{C} \cup \mathcal{X} \rceil_1$ is obtained from $\lfloor \mathcal{C} \cup \mathcal{X} \rceil_0$ by BLK on $\lfloor \mathcal{X} \rceil_1$ and $(\mathcal{F} \perp\!\!\!\perp_0 \mathcal{X} \cup \mathcal{H}) \,|\, \mathcal{G}$ then $(\mathcal{F} \perp\!\!\!\perp_1 \mathcal{H}) \,|\, \mathcal{G}$
3. If $(\mathcal{X} \perp\!\!\!\perp_0 \mathcal{Y}) \,|\, \mathcal{A}$ and $\lfloor \mathcal{B}^* \rceil_{12}$ is obtained by commutative BLK on $\lfloor \mathcal{X} \rceil_1$ and $\lfloor \mathcal{Y} \rceil_2$ then $\lfloor \mathcal{A} \rceil_{12}$ is obtained by commutative BLK on $\lfloor \mathcal{A} \rceil_1$ and $\lfloor \mathcal{A} \rceil_2$

Use these results to generate local computation algorithms.

Conditional Bayes linear Bayes model

- let $\mathcal{X} = \{X_1, \ldots, X_n\}$ be a partition with a probability distribution attached to it. Let $\mathcal{A}$ and $\mathcal{B}$ two collections of random quantities
- judge that given $\mathcal{X}$ we can learn nothing about $\mathcal{A}$ from $\mathcal{B}$ and vice versa
- if, given $\mathcal{X}$, the graphical model over $\mathcal{A}$ is a Bayesian graphical model and the graphical model over $\mathcal{B}$ is a Bayes linear Bayes graphical model then the graphical model over $\mathcal{A} \cup \mathcal{B} \cup \mathcal{X}$ is a conditional Bayes linear Bayes graphical model

Example of a conditional Bayes linear Bayes model



Updating a conditional Bayes linear Bayes model

1. multiply each node in $\mathcal{B}$ by $\mathcal{X}$ and use Bayes linear kinematics to update over these nodes and $\mathcal{X}$
2. in particular, this will give a new distribution over $\mathcal{X}$
3. propagate updates to $\mathcal{A}$ using probability kinematics on $\mathcal{X}$

Some conclusions

- Bayes linear kinematics offer the natural generalisation to probability kinematics for propagating generalised changes in belief through partially specified belief systems, based around expectation as primitive.
- The application of Bayes linear kinematics to Bayes linear Bayes graphical models shows the flexibility that we may achieve with such an approach as it allows us to combine the simplicity and tractability of belief specification and analysis which we usually associate with Gaussian graphical models with the ability to absorb data into the model by conditioning on marginal distributions of any form.