Learning about large industrial systems

David Randell
Philip Jonathan
(Michael Goldstein)

25 February 2010
Introduction and motivation
   Large systems
   Modelling

Updating beliefs
   The Bayes linear approach
   Exchangeable events
   Making decisions

Application: corrosion monitoring
   Corrosion monitoring
   Data characteristics
   Bayes linear variance learning
   Model diagnostics

Conclusions and future work
Large systems

- Research: galaxy evolution, climate change
- Manufacturing: fouling, corrosion, fatigue
- Environmental: ground, water and airborne monitoring
- Commerce: financial, transactional, software
System characteristics

- **High** dimensional (> 1000 variables)
- **Dependent** variables (e.g. in time or space)
- **Evolves** (e.g. in time)
- Observed **with error**
- Observing complete system prohibitively costly
Method components

1. Specify **model**
   - Partial belief structure
   - Exchangeability assumptions (if any)

2. Simulate to estimate **full belief** structure

3. **Adjust expectations** given beliefs and observations
   - Incomplete and irregular observations
   - Learn about system level and (co-)variance structure

4. Simulate adjusted system to **forecast**

5. Make **decision**
   - Expected loss to optimise decision
Typical model specification

- Two spatial dimensions ($l, c$), one temporal ($t$)
- Observations in time ($t$) and **one** spatial dimension ($c$) only
- Observations with error ($\epsilon_{Y_{lct}}$)
- **Global** evolution ($\epsilon_{\Theta_{ct}}$) with respect to $t$ and $c$
- **Local** evolution in $l$ dimension ($\epsilon_{rlct}$) **relative** to global
Typical model form

Observation: \[ Y_{ct} = f_l (Z_{lct} + \epsilon_{Y_{lct}}) \]

System: \[ Z_{lct} = F\Theta_{ct} + r_{lct} \]

Global Effects: \[ \Theta_{ct} = G\Theta_{ct-1} + \epsilon_{\Theta_{ct}} \]

Local Effects: \[ r_{lct} = g(r_{lct-1}) + \epsilon_{rlct} \]

- \( f_l \) reduces (or “integrates” over) \( l \)
- \( g \) describes local evolution
- \( F \) and \( G \) are regression and system evolution matrices
Partial to full beliefs

Specify **partial** beliefs:

- Specify model form $f_l$, $F$, $G$ and $g$
- Specify variance structures $\sigma^2_Y$, $\Sigma_\Theta$ and $\sigma^2_{rl}$
- Specify initial values for $\Theta_{c0}$ and $r_{lc0}$

Estimate **full** beliefs:

- Generate multiple realisations of model evolution
- Calculate empirical estimates for any expectations and (co-)variance structures of interest
  - In particular: $E(Y)$, $\text{Var}(Y)$, $\text{Cov}(Y, \Theta)$
  - Also: $E(\Theta)$, $\text{Var}(\Theta)$ ...
The Bayes linear approach

Full Bayesian modelling of **large systems**:
- Difficult or impractical to make full prior specifications
- Non-physical simplifications required for modelling

Bayes linear modelling:
- Requires specification of **partial beliefs** only
- Is computationally efficient for **high dimensional** problems
- Uses **expectation** as a primitive rather than probability
- Beliefs are updated using **adjusted expectations**
- de Finetti [1974] or Goldstein and Wooff [2007]
Adjusting beliefs

Observe data $D$ to update beliefs $B$

The **adjusted expectation** vector for $B$ given $D$ is:

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

The **adjusted variance** matrix for $B$ given $D$ is:

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

- $E_D(B)$ used as an **updated estimator** for $B$
- $\text{Var}_D(B)$ can be viewed as the **mean square error** of the estimator $E_D(B)$
Motivating Bayes linear

Two collections of random quantities, \( B = (B_1 \ldots B_r) \) and \( D = (D_1 \ldots D_s) \).
The adjusted expectation for \( B_i \) given \( D \) is the linear combination \( a_i^T D \),

\[
E_D(B) = \sum_{i=0}^{s} a_i^T D_i
\]

which minimises;

\[
E \left( (B_i - \sum_{i=0}^{k} a_i^T D_i)^2 \right)
\]

over choices of \( a_i^T \).

- Must specify prior mean vectors and variance matrices for \( B \) and \( D \) and a covariance matrix between \( B \) and \( D \).
Exchangeable events

- In an **exchangeable** sequence of random variables, future samples behave like earlier ones.

- A collection of quantities \( X = \{X_1, X_2, \ldots \} \) is exchangeable if our beliefs are **invariant under permutation** of \( X \).

- The role of exchangeability in subjective analysis is **analogous to that of independence** in classical inference.

- An exchangeable sequence can be represented as a mixture of underlying i.i.d. sequences (de Finetti [1974]).
Independent events are exchangeable, but exchangeable events may not be independent

- A sequence of i.i.d. random variables is exchangeable
- Sampling without replacement is exchangeable, but not independent
- For the bivariate normal random variable:

\[
Z \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right)
\]

components \(Z_1\) and \(Z_2\) are exchangeable, but independent only if \(\rho = 0\)
Second order exchangeability

A collection $X = \{X_1, X_2, \ldots \}$ is second order exchangeable if our beliefs about first and second order specification are invariant under permutation of $X$

$$E(X_i) = \mu \quad \text{Var}(X_i) = \sigma \quad \text{Cov}(X_i, X_j) = \gamma \quad i \neq j$$

- Equivalent to full exchangeability for Bayes linear modelling
The representation theorem

For (s.o.) exchangeable $X = X_1, X_2, \ldots$, we represent each $X_i$ as the sum of two random quantities, a “mean” plus “residual”:

$$X_i = \mathcal{M} + \mathcal{R}_i$$

Each pair $\mathcal{R}_i$ and $\mathcal{R}_j$ are uncorrelated $i \neq j$ and each $\mathcal{R}_i$ is uncorrelated with $\mathcal{M}$ (Goldstein [1986])

$$E(\mathcal{M}) = \mu \quad \text{Var}(\mathcal{M}) = \gamma$$
$$E(\mathcal{R}_i) = 0 \quad \text{Var}(\mathcal{R}_i) = \sigma - \gamma$$

- Simplifies specification of (co-)variance structures
- Adjust beliefs about $\mathcal{M}$ not $X_i$
Exchangeable errors: simple (co-)variance structures

Global Effects: \[ \Theta_{ct} = G\Theta_{ct-1} + \epsilon_{\Theta t} \quad \text{Var}(\epsilon_{\Theta t}) = \Sigma_{\Theta} \]

Assume (s.o.) exchangeability of \( \epsilon_{\Theta ct} \) over \( c \) and \( t \)

\[ \epsilon_{\Theta ct} = M_{\Theta} + R_{\Theta ct} \]

- Then \( \text{Var}(\epsilon_{\Theta ct}) = \sigma^2_{\Theta} \), for all \( c \) and \( t \)
- And \( \text{Cov}(\epsilon_{\Theta c't'}, \epsilon_{\Theta ct}) = \gamma_{\Theta} \), for all \( c' \neq c \) and \( t' \neq t \)
- Hence, a simple **two parameter form** for \( \Sigma_{\Theta} = \Sigma_{\Theta}(\sigma^2_{\Theta}, \gamma_{\Theta}) \)
Exchangeable *squared* errors: (co-)variance learning

Global Effects: \[ \Theta_{ct} = G\Theta_{ct-1} + \epsilon_{\Theta ct} \]
\[ \text{Var}(\epsilon_{\Theta t}) = \Sigma_{\Theta} \]

Assume (s.o.) exchangeability of \( \epsilon^2_{\Theta ct} \) over \( c \) and \( t \)

\[ \epsilon^2_{\Theta ct} = \mathcal{M_V} + \mathcal{R}_{Vct} \]

- Then \( E(\epsilon^2_{\Theta ct}) = E(\mathcal{M_V}) = \sigma^2_{\Theta} \), for all \( c \) and \( t \)
- Hence adjusting beliefs about \( \mathcal{M_V} \) allows us to **learn about** variances
Method components revisited

1. Specify model
   ▶ Partial belief structure
   ▶ Exchangeability assumptions (if any)
2. Simulate to estimate full belief structure
3. Adjust expectations given beliefs and observations
   ▶ Incomplete and irregular observations
   ▶ Learn about system level and (co-)variance structure
4. Simulate adjusted system to forecast
5. Make decision
   ▶ Expected loss to optimise decision
Making decisions: Optimal inspection design

- Identify **good inspection designs** with which to update our beliefs
- Potential designs evaluated in terms of **reducing uncertainty** about **critical system characteristics**
- **Utility** or **loss** is used to compare designs

For example:
- Simple decision to replace or retain a system component subject to potential **costly failure**
Loss for component replacement

- Simple maintenance decision $\delta \in \Delta$ to replace $R$ or retain $\bar{R}$.
- Outcome $o \in O$ is either failure $F$ or survival $\bar{F}$.
- Loss $L(o, \delta)$ is specified as:

\[
\begin{array}{c|cc}
   & F & \bar{F} \\
\hline
R & L_R & L_R \\
\bar{R} & L_F & 0
\end{array}
\]
Expected loss with observed data

For observed data $D$:

$$E_{O|D}[L(O, \delta)|D] = L(F, \delta)\Pr(F|D) + L(\bar{F}, \delta)\Pr(\bar{F}|D)$$

$$E[L(O, R)|D] = L(F, R)\Pr(F|D) + L(\bar{F}, R)\Pr(\bar{F}|D) = LR$$

$$E[L(O, \bar{R})|D] = L(F, \bar{R})\Pr(F|D) + L(\bar{F}, \bar{R})\Pr(\bar{F}|D) = LF\Pr(F|D)$$

Replacement is selected when:

$$E[L(O, R)|D] < E[L(O, \bar{R})|D]$$

$$\Pr(F|D) > \frac{LR}{LF}$$
Expected loss with unobserved data

Expected loss of decision $\delta$ based on as yet unobserved data $D$ from design $d$ is:

$$E_{\mathcal{O}}[L(O, \delta)] = E_{\mathcal{D}}\{E_{\mathcal{O}|\mathcal{D}}[L(O, \delta)|D]\}$$
$$= E_{\mathcal{D}}\{L(F, \delta)Pr(F|D) + L(\bar{F}, \delta)Pr(\bar{F}|D)\}$$

Optimal decision $\delta^*$ satisfies:

$$\delta^* = \begin{cases} 
R & \text{if } Pr(F|D) > \rho \\
\bar{R} & \text{if } Pr(F|D) \leq \rho 
\end{cases}$$

where $\rho = \frac{L_R}{L_F}$
Expected loss for design, \( E[L(O, \delta^*)] \)

\[
E_O[L(O, \delta^*)] = E_D\{E_{O|D}[L(O, \delta^*)| D]\}
= E_D\{L(F, \delta^*)Pr(F|D) + L(\bar{F}, \delta^*)Pr(\bar{F}|D)\}
= E\{L(F, \delta^*)Pr(F|D) + L(\bar{F}, \delta^*)Pr(\bar{F}|D)|\delta^* = R\}Pr(\delta^* = R) + E\{L(F, \delta^*)Pr(F|D) + L(\bar{F}, \delta^*)Pr(\bar{F}|D)|\delta^* = \bar{R}\}Pr(\delta^* = \bar{R})
= LRPr(Pr(F|D) > \rho) + LF E\{Pr(Pr(F|D)|Pr(Pr(F|D) \leq \rho)\}Pr(Pr(F|D) \leq \rho)
= LRl_1 + LF l_2
Expected loss for **design**, $E[L(O, \delta^*)]$ 

$$E[L(O, \delta^*)] = L_R l_1 + L_F l_2$$

- Integrals $l_1$ and $l_2$ evaluated for **given** probability distributions characterised by **location** and **scale** parameters
- Adjusted expectations and variances from the Bayes linear update used to estimate location and scale
- Computationally fast: **no need to simulate** data $D$ for given design $d$
Application: Corrosion monitoring of offshore platform
Corrosion monitoring

- Offshore platforms have large numbers of components subject to corrosion
- Corrosion can lead to failure incurring costs
- A typical offshore platform has $>100$ corrosion circuits, each with 20 to 1000 components, hence potentially $>5000$ components subject to corrosion.
- Some corrosion circuits have similar characteristics
Learning about large industrial systems

- Application: corrosion monitoring
- Corrosion monitoring

Typical corrosion circuit diagram
Data characteristics

- **Minima**: over whole component observed
- **Short time series**: data per component is limited, but large number of components
- **Irregular inspections**: inspections are carried out when possible, often when processes are shut down, often several months or years apart
- **Incomplete inspections**: due to size of systems and inaccessibility of components, complete systems are rarely inspected
Typical inspection design for a corrosion circuit
Method components

1. Specify model
   - Partial belief structure
   - Exchangeability assumptions (if any)
2. Simulate to estimate full belief structure
3. Adjust expectations given beliefs and observations
   - Incomplete and irregular observations
   - Learn about system level and (co-)variance structure
4. Simulate adjusted system to forecast
5. Make decision
   - Expected loss to optimise decision
The system is modelled as:

\[ Y_{tc} = \min_{l} (X_{tc} + r_{tcl} + \epsilon_{Y_{tcl}}) \]

\[ X_{tc} = X_{t-1c} + \alpha_{tc} + \epsilon_{X_{tc}} \]

\[ \alpha_{tc} = \alpha_{t-1c} + \epsilon_{\alpha_{tc}} \]

\[ r_{tcl} = r_{t-1cl} + \epsilon_{r_{tcl}} \]

\[ \text{Var}(\epsilon_{Y_{tcl}}) = \sigma_{Yc}^2 \]

\[ \text{Var}(\epsilon_{X_{tc}}) = \Sigma_X \]

\[ \text{Var}(\epsilon_{\alpha_{tc}}) = \Sigma_\alpha \]

\[ \text{Var}(\epsilon_{r_{tcl}}) = \sigma_{rc}^2 \]
Learning about large industrial systems

Application: corrosion monitoring

Data characteristics

Learning about wall thickness and corrosion rate

- Perform simulations of model based on partial belief specification
- Simulations together with inspection data yield updated adjusted expectations for wall thickness and corrosion rate parameters
- Modelling covariance structure, we learn about all components even unobserved
Typical covariance structure based on adjacency
Variance learning: why?

- Prior specification of (co-)variances is difficult.
- Variance parameters in model typically fixed. Poor prior specification leads to poor model performance.
- Variance is not directly observable. Adjusting beliefs more difficult.
Variance learning: simple corrosion model

For example:

\[ X_{ct} = X_{ct-1} + \alpha_{ct} + \epsilon_{Xct} \]
\[ \alpha_{ct} = \alpha_{ct-1} + \epsilon_{\alpha ct} \]

Differences of observations eliminate effects of wall thickness’
and corrosion rates (Wilkinson [1997])

\[ X_t^{(1)} = X_{ct} - X_{ct-1} = \alpha_{ct} + \epsilon_{Xct} = \alpha_{ct-1} + \epsilon_{\alpha ct} + \epsilon_{Xct} \]
\[ X_t^{(2)} = X_{ct} - X_{ct-2} = X_{ct-1} + \alpha_{ct} - X_{ct-2} + \epsilon_{Xct} \]
\[ = \alpha_{ct} + \alpha_{ct-1} + \epsilon_{Xct} + \epsilon_{Xct-1} \]
\[ = 2\alpha_{ct-1} + \epsilon_{\alpha ct} + \epsilon_{Xct} + \epsilon_{Xct-1} \]
Variance learning: squared differences

Therefore:

$$X_t^{(2)} - 2X_t^{(1)} = -\epsilon \alpha c t - \epsilon X c t + \epsilon X c t - 1$$

and:

$$E[(X_t^{(2)} - 2X_t^{(1)})^2] = E[(-\epsilon \alpha c t - \epsilon X c t + \epsilon X c t - 1)^2]$$

$$= E[\epsilon^2_{\alpha c t}] + E[\epsilon^2_{X c t}] + E[\epsilon^2_{X c t - 1}]$$

$$= \sigma^2_{\alpha c} + 2\sigma^2_{X c}$$
Variance learning: exchangeability in time

Assume squares of residuals are (s.o.) exchangeable in time. Using representation theorem:

\[ [\epsilon_{X_{ct}}]^2 = \mathcal{M}(V_c) + \mathcal{R}_t(V_c) \]

where:

\[ E([\epsilon_{X_{ct}}]^2) = \sigma^2_{X_c} = V_c \quad \text{Var}([\epsilon_{X_{ct}}]^2) = \Sigma_{V_c} \]
\[ \text{Cov}([\epsilon_{X_{ct}}]^2, [\epsilon_{X_{ct'}}]^2) = \Gamma_{V_c} \quad t \neq t' \]
Variance learning: adjusting beliefs

Compute $E_D[M(V_c)]$:

$$D = \left\{ \frac{(X_{t}^{(2)} - 2X_{t}^{(1)})^2}{2 + \lambda} \right\}^T_{t=3}$$

$$E_D[M(V_c)] = E[M(V_c)] + \text{Cov}[M(V_c), D]\text{Var}[D]^{-1}(D - E(D))$$

$$= \sigma^2_{\hat{X}_c} + 2'_{T} \Gamma_{V_c} \text{Var}[D]^{-1}(D - 1_T(\sigma^2_{\alpha_c} + 2\sigma^2_{\hat{X}_c}))$$

yielding an adjusted estimate for the variances in the model.
Variance learning: generalisations

Generalisations include:

- **General time step** form for irregular time points
- Partial inspections using *exchangeable variances* across components
- **Mahalanobis distance fitting** to update local variances
Model diagnostics

- Mahalanobis distance to estimate data discrepancy, comparing data to our prior estimates

\[
\text{Dis}(X) = \frac{(D - E(D))^2}{\text{Var}D}
\]

- For each of our updated values we can also compute the adjustment discrepancy

\[
\text{Dis}_D(X) = \frac{(E_D(X) - E(X))^2}{R\text{Var}_D X}
\]
Typical model diagnostics

![Graph showing discrepancy over time](image-url)
Conclusions

General purpose framework for modelling and inspection design of large systems

Compared to existing methods, the model is novel in that:

- Analysis of multivariate systems possible, rather than modelling components separately and independently
- Data from incomplete inspections at arbitrary times used to learn about the whole system
- Uncertainties in system parameters adjusted, as are the dependencies between these
- Economically-optimal future inspection strategies can be estimated consistently
Future work

- Efficient implementation of **sequential** Bayes linear calculation
- **Search methods** for good designs in high dimensions
- **Elicitation** of prior partial beliefs
- Flexible forms for modelling for system element behaviour
- Enhanced criteria for **evaluation of inspection schemes**
- **Fundamental modelling** of physical processes (e.g. corrosion)
- **New applications** to manufacturing, environmental and commercial problems
Thank you

david.randell@durham.ac.uk
philip.jonathan@shell.com

Randell et al. [February 2010]

www.lancaster.ac.uk/~jonathan


Backup

\[ E_Y(g(Y)) = E_X(E_{Y|X}(g(Y)|X)) \]
\[ E_{Y|X}(g(Y)|X) = \sum_i g(Y_i)\Pr(Y_i|X) \]