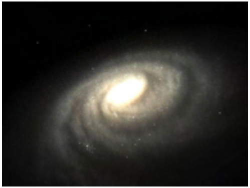
Learning about large industrial systems

David Randell
Philip Jonathan
(Michael Goldstein)

25 February 2010

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 (ϵ_{Ylct})
- **Global** evolution ($\epsilon_{\mathbf{\Theta ct}}$) with respect to t and c
- **Local** evolution in l dimension (ϵ_{rlct}) **relative** to global

Typical model form

$$
\begin{aligned}
\text{Observation:} \quad & Y_{ct} = f_l\left(Z_{lct} + \epsilon_{Ylct}\right) & \mathrm{Var}(\epsilon_{Ylct}) = \sigma_Y^2 \\
\text{System:} \quad & Z_{lct} = \mathbf{F}\mathbf{\Theta}_{\mathbf{ct}} + r_{lct} & \\
\text{Global Effects:} \quad & \mathbf{\Theta}_{\mathbf{ct}} = \mathbf{G}\mathbf{\Theta}_{\mathbf{ct-1}} + \epsilon_{\mathbf{\Theta ct}} & \mathrm{Var}(\epsilon_{\mathbf{\Theta t}}) = \Sigma_{\Theta} \\
\text{Local Effects:} \quad & r_{lct} = g(r_{lct-1}) + \epsilon_{rlct} & \mathrm{Var}(\epsilon_{rlct}) = \sigma_{rl}^2
\end{aligned}
$$

- 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, $\mathbf{F}$, $\mathbf{G}$ and g
- Specify variance structures σ_Y^2, Σ_Θ and σ_{rl}^2
- Specify initial values for $\mathbf{\Theta_{c0}}$ 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: $E(\mathbf{Y})$, $\mathrm{Var}(\mathbf{Y})$, $\mathrm{Cov}(\mathbf{Y}, \mathbf{\Theta})$
  - Also: $E(\mathbf{\Theta})$, $\mathrm{Var}(\mathbf{\Theta})$ ...

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:

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

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

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

- $E_D(B)$ used as an **updated estimator** for B
- $\mathrm{Var}_D(B)$ 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])

Exchangeability and independence

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

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

- 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])

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

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

Exchangeable errors: simple (co-)variance structures

Global Effects: $\mathbf{\Theta_{ct}} = \mathbf{G}\mathbf{\Theta_{ct-1}} + \epsilon_{\mathbf{\Theta ct}} \qquad \mathrm{Var}(\epsilon_{\mathbf{\Theta t}}) = \Sigma_{\Theta}$

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

$$\epsilon_{\Theta ct} = \mathcal{M}_{\Theta} + \mathcal{R}_{\Theta ct}$$

- Then $\mathrm{Var}(\epsilon_{\Theta ct}) = \sigma^2_{\Theta}$, for all c and t
- And $\mathrm{Cov}(\epsilon_{\Theta c't'}, \epsilon_{\Theta ct}) = \gamma_{\Theta}$, 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: $\mathbf{\Theta_{ct}} = \mathbf{G}\mathbf{\Theta_{ct-1}} + \epsilon_{\mathbf{\Theta ct}} \qquad \mathrm{Var}(\epsilon_{\mathbf{\Theta t}}) = \Sigma_{\Theta}$

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:

	F	$\bar{F}$
R	L_R	L_R
$\bar{R}$	L_F	0

Expected loss with observed data

For **observed** data D:

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

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

Replacement is selected when:

$$E[L(O,R)|D] < E[L(O,\bar{R})|D]$$
$$\mathrm{Pr}(F|D) > \frac{L_R}{L_F}$$

Expected loss with **unobserved** data

Expected loss of decision δ based on **as yet unobserved** data D from **design** d is:

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

Optimal decision δ^* satisfies:

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

Expected loss for **design**, $E[L(O,\delta^*)]$

$$E_{[O]}[L(O,\delta^*)]$$

$$
\begin{aligned}
&= E_{[D]}\{E_{[O|D]}[L(O,\delta^*)|D]\}\\
&= E_{[D]}\{L(F,\delta^*)\Pr(F|D) + L(\bar{F},\delta^*)\Pr(\bar{F}|D)\}\\
&= E\{L(F,\delta^*)\Pr(F|D) + L(\bar{F},\delta^*)\Pr(\bar{F}|D)|\delta^* = R\}\Pr(\delta^* = R)\\
&+ E\{L(F,\delta^*)\Pr(F|D) + L(\bar{F},\delta^*)\Pr(\bar{F}|D)|\delta^* = \bar{R}\}\Pr(\delta^* = \bar{R})\\
&= L_R\Pr(\Pr(F|D) > \rho)\\
&+ L_F E\{\Pr(\Pr(F|D)|\Pr(\Pr(F|D) \le \rho\}\Pr(\Pr(F|D) \le \rho)\\
&= L_R I_1 + L_F I_2
\end{aligned}
$$

Expected loss for **design**, $E[L(O, \delta^*)]$

$$E[L(O, \delta^*)] = L_R I_1 + L_F I_2$$

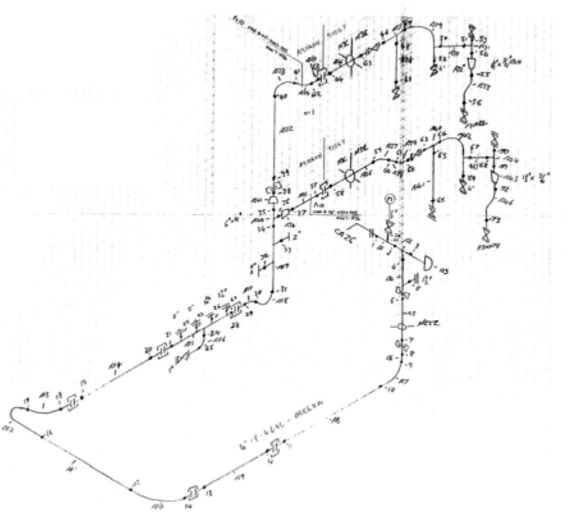
- Integrals I_1 and I_2 evaluated for **given** probability distributions characterised by **location** and **scale** parameters
- Adjusted expectations and variances from the Bayes linear update used to estimate location and scale
- Computationally fast: **no need to simulate** data D for given design 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

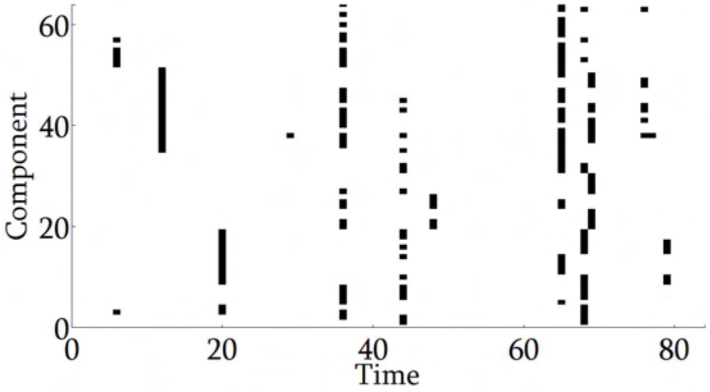
Typical corrosion circuit diagram



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

Model

The system is modelled as:

$$Y_{tc} = \min_l \left(X_{tc} + r_{tcl} + \epsilon_{Ytcl}\right) \qquad \mathrm{Var}(\epsilon_{Ytcl}) = \sigma^2_{Yc}$$

$$X_{tc} = X_{t-1c} + \alpha_{tc} + \epsilon_{Xtc} \qquad \mathrm{Var}(\epsilon_{Xtc}) = \Sigma_X$$

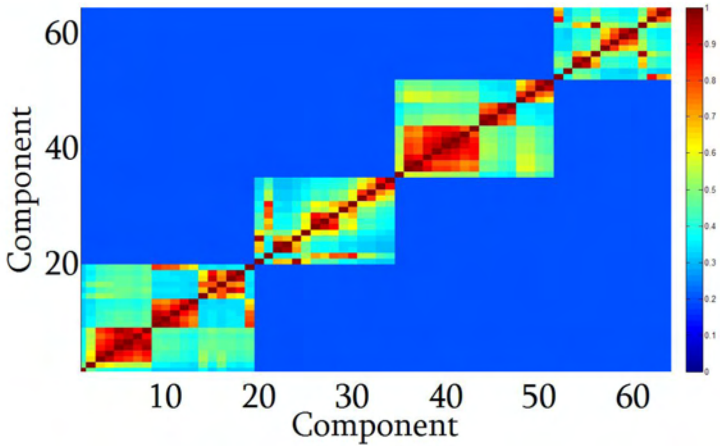
$$\alpha_{tc} = \alpha_{t-1c} + \epsilon_{\alpha tc} \qquad \mathrm{Var}(\epsilon_{\alpha tc}) = \Sigma_\alpha$$

$$r_{tcl} = r_{t-1cl} + \epsilon_{rtcl} \qquad \mathrm{Var}(\epsilon_{rtcl}) = \sigma^2_{rc}$$

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])

$$\begin{aligned} X_t^{(1)} = X_{ct} - X_{ct-1} &= \alpha_{ct} + \epsilon_{Xct} = \alpha_{ct-1} + \epsilon_{\alpha ct} + \epsilon_{Xct} \\ X_t^{(2)} = X_{ct} - X_{ct-2} &= X_{ct-1} + \alpha_{ct} - X_{ct-2} + \epsilon_{Xct} \\ &= \alpha_{ct} + \alpha_{ct-1} + \epsilon_{Xct} + \epsilon_{Xct-1} \\ &= 2\alpha_{ct-1} + \epsilon_{\alpha ct} + \epsilon_{Xct} + \epsilon_{Xct-1} \end{aligned}$$

Variance learning: squared differences

Therefore:

$$X_t^{(2)} - 2X_t^{(1)} = -\epsilon_{\alpha ct} - \epsilon_{Xct} + \epsilon_{Xct-1}$$

and:

$$\begin{aligned} E[(X_t^{(2)} - 2X_t^{(1)})^2] &= E[(-\epsilon_{\alpha ct} - \epsilon_{Xct} + \epsilon_{Xct-1})^2] \\ &= E[\epsilon_{\alpha ct}^2] + E[\epsilon_{Xct}^2] + E[\epsilon_{Xct-1}^2] \\ &= \sigma_{\alpha c}^2 + 2\sigma_{Xc}^2 \end{aligned}$$

Variance learning: exchangeability in time

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

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

where:

$$E([\epsilon_{Xct}]^2) = \sigma^2_{Xc} = V_c \quad \mathrm{Var}([\epsilon_{Xct}]^2) = \Sigma_{V_c}$$

$$\mathrm{Cov}([\epsilon_{Xct}]^2, [\epsilon_{Xct'}]^2) = \Gamma_{V_c} \quad t \neq t'$$

Variance learning: adjusting beliefs

Compute $E_D[\mathcal{M}(V_c)]$:

$$D = \left\{ \frac{(X_t^{(2)} - 2X_t^{(1)})^2}{2+\lambda} \right\}_{t=3}^{T}$$

$$\begin{aligned} E_D[\mathcal{M}(V_c)] &= E[\mathcal{M}(V_c)] + \mathrm{Cov}[\mathcal{M}(V_c), D]\mathrm{Var}[D]^{-1}(D - E(D)) \\ &= \sigma^2_{Xc} + 2'_T \Gamma_{V_c} \mathrm{Var}[D]^{-1}(D - 1_T(\sigma^2_{\alpha c} + 2\sigma^2_{Xc})) \end{aligned}$$

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

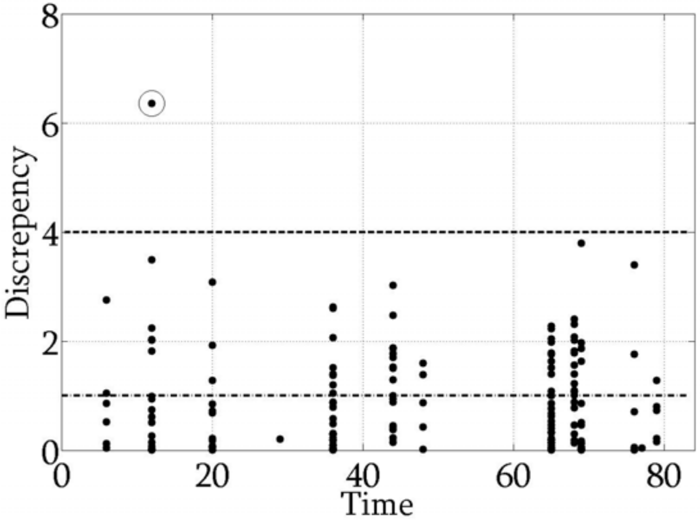
- Mahalanobis distance to estimate **data** discrepancy, comparing data to our prior estimates

$$\mathrm{Dis}(X) = \frac{(D - E(D))^2}{\mathrm{Var}D}$$

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

$$\mathrm{Dis}_D(X) = \frac{(E_D(X) - E(X))^2}{\mathrm{RVar}_D X}$$

Typical model diagnostics



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

B. de Finetti. *Theory of Probability*. Wiley series in probability and mathematical statistics. Wiley, London, 1974.

M. Goldstein. Exchangeable belief structures. *Journal of the American Statistical Association*, 81:971–976, 1986.

M. Goldstein and D. Wooff. *Bayes Linear Statistics: Theory and Methods*. Wiley Series in Probability and Statistics. Wiley, New York USA, 2007.

D. Randell, M. Goldstein, G. Hardman, and P. Jonathan. Bayesian linear inspection planning for large scale physical systems. *Submitted to Journal of Risk and Reliability*, February 2010.

D. J. Wilkinson. Bayes linear variance adjustment for locally linear dlms. *Journal of Forecasting*, 16:329–342, 1997.

Backup

$$E_{[Y]}(g(Y)) = E_{[X]}(E_{[Y|X]}(g(Y)|X))$$

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