Bayes Linear Bayes Graphical Models in the Design of Optimal Test Strategies

KEVIN J. WILSON, JOHN QUIGLEY, TIM BEDFORD and LESLEY WALLS
Department of Management Science, University of Strathclyde, U.K.

(Received on April 10, 2013, revised on May 08, and June 11, 2013)

Abstract: Test and analysis plays a vital role in reducing uncertainty about the true performance of an engineering system. However tests can be expensive and designing an optimal test strategy can be challenging. We propose a Bayesian modelling process, which takes the form of a Bayesian Network, to determine anticipated test efficacy. Such a model supports engineering managers in assessing trade-offs between test resources and uncertainty reduction. Inference based on a full Bayesian model can be computationally demanding to the extent that it can limit practical application. To overcome this constraint, we develop a Bayes linear approximation for inference. This approach is known as a Bayes linear Bayes graphical model. After explaining the key principles of the method, we provide an application to a real industrial test to establish the condition of an ageing engineering system.

Keywords: Bayesian network, Bayes linear kinematics, engineering test strategy, reliability analysis, uncertainty quantification.

1. Introduction

Tests provide a means of assessing the true state of an engineering system in terms of reliability performance, condition of assets, etc. Without test and analysis we would have considerable uncertainty about true system performance. Designing a perfect test is the ultimate goal, but will not usually be met. The best we may achieve is to “buy-down” uncertainty by gaining the most useful information about the system. The term “system” may be replaced by sub-assembly, component, part or item as appropriate.

This work is driven by tests related to understanding reliability performance. This may be in relation to design of a development test and analysis strategy to reduce uncertainty about anticipated weaknesses or potential failure modes and causes [1, 2], or to understanding of condition of an ageing system to support assessment of uncertainty in remaining life [3]. The method developed need not be restricted to reliability tests.

We utilise a graphical model considering probabilistic relationships between true performance of a system and that observed on test. A Bayesian Network (BN) [4] is developed to represent the reasoning of engineers who specify variables defining performance characteristics, identify causal structure and provide subjective probability assessments. The model anticipates parameter uncertainties following each test and supports the identification of the optimal test, known as pre-posterior analysis.

A modern challenge of using a BN in our context is computational effort. Models consist of many variables, each with multiple states and dependency, which cannot be solved analytically. Numerical or simulation based methods are used to evaluate the posterior distribution such as Markov Chain Monte Carlo (MCMC), which is computationally intensive for problems with many parameters [5].

We develop a Bayes linear [6] approach to support inference which approximates the full Bayesian analysis. Beliefs are adjusted by linear fitting and results in posterior
expectations, variances and covariances. All updates are made analytically. The most demanding computation is the numerical inversion of a covariance matrix. The choice of an optimal test can be made using Bayes linear methods when the full Bayesian solution is computationally impossible.

Bayes linear methods reduce the elicitation burden on experts from a multivariate probability distribution to first and second order moments. Humans are poor at assessing moments directly [7] so priors can be specified using indirect methods such as quantiles [8]. Bayes linear provides a good approximation to full Bayes for models which are approximately linear. In some cases the approximation is exact. The inclusion of higher order moments can improve approximation in non-linear models.

A natural way to define the optimal test is the one which brings about the largest gain in information, or reduction in uncertainty, about some underlying state of system performance. We choose the test prior to observing outcomes and so optimise the expectation of the gain in information. We consider an optimality criterion based on maximising the expected reduction in the variance between prior and posterior.

We consider situations where test observations can be classified into a finite number of discrete states. For example, a visual inspection of a system might indicate the presence or absence of a defect. If we are measuring continuous variables, such as the amount of trace elements in engine oil, then we can categorise data into discrete states, though we lose information in this transformation process. We suppose that we are to make a single set of observations from a multinomial distribution.

The article is structured as follows. Section 2 considers Bayes linear methods, Bayes linear Bayes graphical models, conjugate Bayesian updating for multinomial distributions and specifying optimal test strategies. In Section 3 we consider an industrial application for an ageing system. We conclude with a review of our proposed approach and ideas for further work in Section 4.

2. Bayesian Networks and Bayes linear Bayes Graphical Models

2.1 Bayesian Networks

A BN [9, 4] is a collection of nodes and arcs that form a Directed Acyclic Graph. Associated with each arc is a probability distribution measuring the association between variables represented as nodes on either end of the arc. The direction of the arc represents the causal direction of the relationship.

Consider the simple example of part of a BN given for 3 generic engineering variables in Figure 1.

The relationship illustrated in Figure 1 indicates that Pitting and Surface Corrosion are not associated directly, but both influence the presence of Cracks. We refer to such a situation between Surface Corrosion and Pitting as having converging connections. As we increase the number of arcs that converge on a node, the number of probabilities which require to be estimated grows geometrically. For example, for the BN in Figure 1 we would require four probability distributions to be estimated; one for each combination of the conditioning states. In large BNs this results in issues around computation time and effort when carrying out analyses.
We shall use a structure called a Bayes linear Bayes graphical model to reduce this computation time. To do so, we first introduce Bayes linear methods.

2.2 Bayes Linear Methods

Bayes linear methods [6, 10] offer an alternative subjective approach to inference to the full Bayesian approach. They use expectation as their primitive rather than probability. The advantage of this is that to evaluate an expectation requires all probabilities whereas a probability is simply the expectation of a single indicator variable.

Suppose that the variables we shall observe are $M_{g21eM34n1M4PPPM1e3n2e}$ and we use these observations to update beliefs about $M_{g215eM34n1M4PPb}$, variance matrices $V_{g215eM34n1M4PPb}$ and covariance matrix $C_{g215eM34n1M4PPb}$. These values would typically be elicited from engineers. Having observed $M_{g21eM34n1M4PPb}$, the adjusted expectation vector and variance matrix of $B$ are found by linear fitting:

$$E_D(B) = E_0(B) + Cov_0(B, D)V_0^{-1}(D)[D - E_0(D)],$$
$$Var_D(B) = Var_0(B) - Cov_0(B, D)V_0^{-1}(D)Cov_0(D, B).$$

If the inverse $V_0^{-1}(D)$ does not exist then a generalised inverse can be used. In the example this would give us our updated expectation and variance for the amount of pitting on the components a month from now having observed the amount of pitting on the same components a month from now. The prior takes the form of the expectation vectors $E_0(D), E_0(B)$, variance matrices $Var_0(D), Var_0(B)$ and covariance matrix $Cov_0(B, D)$. These values would typically be elicited from engineers. Having observed $D_1, ..., D_n$, the adjusted expectation vector and variance matrix of $B$ are found by linear fitting:

$$E_D(B) = E_0(B) + Cov_0(B, D)V_0^{-1}(D)[D - E_0(D)],$$
$$Var_D(B) = Var_0(B) - Cov_0(B, D)V_0^{-1}(D)Cov_0(D, B).$$

If the inverse $V_0^{-1}(D)$ does not exist then a generalised inverse can be used. In the example this would give us our updated expectation and variance for the amount of pitting on the components a month from now having observed the amount of pitting on the components today. The linear relationships given here are standard relationships for the multivariate normal that are applied to general variables and that are also found by obtaining the best linear prediction to a general variable.

Now suppose that we don’t observe $D$ directly but observe some other quantity $I$ which updates our beliefs to $E(D|I)$ and $Var(D|I)$. This could be by full Bayesian conditioning. We then wish to propagate these changes to $B$.

For example, if $D$ were the probabilities that the amount of pitting in the components was above a certain threshold today and $B$ were the probabilities of exceeding the threshold a month from now then $I$ could be the observation of the pitting in a small area of one of the components today. This would change our beliefs about the probability of being over the threshold for that component today.

Propagating these changes in belief in a full Bayesian model would typically involve intensive procedures such as MCMC. However, in a Bayes linear model, we can find the adjusted expectation vector and variance matrix of $A = D \cup B$ as

![Figure 1: BN Illustrating the Causal Relationship between Surface Corrosion, Pitting and Cracks](image-url)
\[ E_{\theta}(A) = E_0(A) + \text{Cov}_0(A, D)\text{Var}_0^{-1}(D)[E(D|I) - E_0(D)] \]  
(1)

\[ \text{Var}_{\theta}(A) = \text{Var}_0(A) + W_0(A, D)\text{Var}(D|I)W_0^T(A, D), \]  
(2)

where \( W_0(A, D) = \text{Cov}_0(A, D)\text{Var}_0^{-1}(D) \). This is known as Bayes linear kinematics [11]. The result in our example would be our updated expectation and variance of the probabilities of pitting being above the threshold in the different components both today and in a month’s time.

### 2.3 Bayes Linear Bayes Graphical Models

Bayes linear Bayes graphical models [11] are, like BNs, graphs in which nodes represent model parameters, arcs represent relationships between parameters and the direction of arcs specify the child and parent nodes. However, instead of conditional independence relationships, an analogous property known as (second order) belief separation [12] is used. This says that for three variables \( \theta_1, \theta_2, \theta_3 \), \( \theta_2 \) separates \( \theta_1, \theta_3 \) if they are conditionally uncorrelated given adjustment by \( \theta_2 \). A directed acyclic graph is a directed graphical model if all pairs of variables are separated by their parents.

A Bayes linear Bayes graphical model is then defined by the following three conditions:

1. A second order prior specification is made of the expectation vectors and variance matrices of each \( \theta_i \); and a covariance matrix is specified between each pair of parent and child nodes.

2. A full Bayesian probability specification is made for each pair \((\theta_i, I_i)\).

3. Each \( I_i \) is conditionally independent of the rest of the network given \( \theta_i \).

Information moves through the Bayes linear Bayes graphical model using a combination of Bayes Theorem and Bayes linear kinematics. When \( I_i \) is observed, Bayes Theorem is used to find \( E(\theta_i|I_i) \) and \( \text{Var}(\theta_i|I_i) \). These updates are then propagated through to other quantities in the network using the Bayes linear kinematic updating equations, (1) and (2).

We can propagate multiple belief changes through a Bayes linear Bayes graphical model using local computations. If we receive information \( I_1, I_3 \) which updates our beliefs about the expectation vectors and variances matrices of \( \theta_2 \) and if \( \theta_2 \) separates \( \theta_1, \theta_3 \), then

\[ \text{Var}_{(2)}(\theta_2; I_1, I_3) = \{\text{Var}_{\theta_2|I_1}(\theta_2) + \text{Var}_{\theta_3|I_3}(\theta_2) - \text{Var}_{\theta_2}(\theta_2)\}^{-1}, \]

\[ E_{(2)}(\theta_2; I_1, I_3) = \text{Var}_{(2)}^{-1}(\theta_2; I_1, I_3)[\text{Var}_{\theta_1|I_1}(\theta_2)E_{\theta_1|I_1}(\theta_2) + \text{Var}_{\theta_3|I_3}(\theta_2)E_{\theta_3|I_3}(\theta_2) - \text{Var}_{\theta_2}(\theta_2)E_{\theta_2}(\theta_2)]. \]

We must be careful when making multiple updates using Bayes linear kinematics as a unique commutative solution does not always exist [11].

Consider condition 2. Suppose that we are to make a single set of observations from a multinomial distribution. That is,

\[ X|\theta \sim \text{Mn}(n, \theta), \]

where \( X = (X_1, ..., X_n) \) is the vector of responses, \( \theta = (\theta_1, ..., \theta_p) \) is the vector of probabilities that a response falls in each of the categories and \( n \) is the total number of observations. Suppose we were to observe just a single event. Then the conditional
probabilities of observing the event in each category given that it hasn’t been observed in previous categories are given by

\[ \lambda_i = \Pr(X_i = 1|X_i = 0, \ldots, X_{i-1} = 0) \]

Unlike the unconditional probabilities, there is no restriction on the conditional probabilities to sum to one. This is useful both in terms of eliciting information from experts, allowing them to express their knowledge on the probabilities without the risk of being incoherent, and also when defining the form of the prior distribution, as we shall see. Therefore, we can give each of them, except for \( \lambda_p = 1 \), marginal beta distributions,

\[ \lambda_i \sim \text{beta}(a_i, b_i), \quad i = 1, \ldots, p - 1, \]

where \( \lambda_i, \lambda_j \) are independent for \( i \neq j \). Doing this, rather than assuming a full Dirichlet distribution, means that we do not have to make artificial assumptions about the covariances between parameters.

The unconditional probabilities of an observation being realised in a certain category are then

\[ \theta_i = \lambda_i(1 - \lambda_{i-1}) \ldots (1 - \lambda_1). \]

Due to the conjugate structure of the prior and likelihood, having observed \( x \) the posterior distribution is still of beta form. In particular, it is

\[ \lambda_i | x \sim \text{beta}(a_i + x_i, b_i + \sum_{j=1}^{i-1} x_j + p - i) \] (3)

We wish to choose the optimal test with respect to the reduction in uncertainty at a single node. The prior variance of \( \lambda \) can be expressed as

\[ \text{Var}_0(\lambda) = E_x[\text{Var}(\lambda | x_t)] + \text{Var}_x[E(\lambda | x_t)] \] (4)

where \( x_t \) is the observation realised from test \( t \). In this we choose the design which reduces the variance at that node by the largest amount between prior and posterior. That is, we seek \( \min_t \{\text{Var}(\lambda | x_t)\text{Var}_0^{-1}(\lambda)\} \). When choosing which test to carry out we do not yet have data. Instead, we calculate the posterior variance for each possible realisation of the data and minimize the expected change in variance between prior and posterior;

\[ \min_t \{E_x[\text{Var}(\lambda | x_t)\text{Var}_0^{-1}(\lambda)]\} \] (5)

Typically, the posterior distribution cannot be computed analytically and numerical methods such as MCMC are used instead. Integrals of this type may have to be calculated many thousands of times; once over each possible data realisation. Solving (5) quickly becomes impossible unless suitable approximation methods are used.

Instead, consider the equivalent Bayes linear Bayes directed graphical model. We can express the prior variance in a Bayes linear model in terms of the adjusted variance as in (4). That is

\[ \text{Var}_a(\lambda) = \text{Var}_1(\lambda; x_t) + \text{RVar}(\lambda; x_t), \]

where \( \text{RVar}(\lambda; x_t) \) is the resolved variance in \( \lambda \) having observed \( x_t \). The optimal information gain from the tests is

\[ \min_t \{E_x[\text{Var}(\lambda; x_t)\text{Var}_0^{-1}(\lambda)]\}. \] (6)
We see that the definitions (5) and (6) are equivalent except that the full Bayes posterior variance matrix has been replaced by the Bayes linear kinematic adjusted variance matrix. The definitions (5) and (6) are ambiguous, however. As each variance is a matrix then the expectation will be a matrix. The minimum of a matrix can be defined in different ways. To remove this ambiguity we instead express (6) as

$$\min_{\lambda} \{ \mathbb{E}_{\lambda}(trace(\text{Var}_{\lambda}(\lambda;x)) \text{Var}_{\lambda}^{-1}(\lambda)) \}$$

This can be thought of as a Bayes linear equivalent to the Bayesian A-optimality criterion [13, 14]

3 Illustrative Example: Optimal Tests of the Condition of an Ageing System

3.1 Background to the Test Decision Problem

This example is based on a de-sensitised industrial case of an ageing complex system where alternative tests were being considered to measure condition and aspects of performance so that better assessments could be made to assess the remaining useful life and the need for any investment. Many uncertainties existed in relation, for example, to the degree of corrosion, pitting, crack, grouting, loading and other indicators of condition and performance of the component parts of the system.

The alternative tests being considered included on-site inspections, laboratory inspections of samples and pull-off tests. The purpose of the analysis was to identify which test had the greatest anticipated uncertainty reduction in the state of the grout condition in the sampled components, all of which are nominally identical. Cost was not a criterion for this aspect of the project.

Figure 5 shows the BN constructed to represent this situation, which was based on the structured elicitations with four engineering experts.

Figure 2: The BN representing the relationships between the condition of the physical system and the six possible tests.
3.2 The Model

Consider the BN which resulted from the elicitation sessions with engineers in Figure 2. For each possible test we would observe a number of observations in each test category, \( X = (X_1, \ldots, X_p) \) which, conditional on the probabilities of being in each category, \( \theta = (\theta_1, \ldots, \theta_p) \), are multinomially distributed \( X|\theta \sim \text{Mn}(n, \theta) \).

We can define the conditional probabilities of an observation from a test of the sampled components falling in a category given that it is not in a previous category, denoted \( \lambda_i \), and express the unconditional probabilities in terms of these quantities. That is,

\[
\theta_i = \lambda_i (1 - \lambda_{i-1}) \cdots (1 - \lambda_1),
\]

for \( i = 1, \ldots, p \), where \( \lambda_p = 1 \). Now consider an underlying physical state variable over which we wish to gain information by conducting one of the possible tests. For instance, this variable could be Grout Condition. This variable also follows a multinomial distribution \( Y|\mu \sim \text{Mn}(n, \mu) \), where \( \mu \) are the probabilities of being in the various categories (good condition, fair condition and poor condition in the case of Grout Condition). We again define conditional probabilities, this time denoted \( \mu_i \),

\[
\gamma_i = \mu_i (1 - \mu_{i-1}) \cdots (1 - \mu_2).
\]

As the values do not depend on any other variables, in the case of Grout Condition, in the BN, we can define prior distributions on the conditional probabilities directly. Suppose they take the form of beta distributions

\[
\mu_i \sim \text{beta}(\alpha_i, \beta_i),
\]

for hyper-parameters \( \alpha_i \) and \( \beta_i \).

Now, we can use the structure of the BN in order to express the probabilities of each of the possible test results in terms of how they come about and, in particular, in terms of the probabilities of each of the states of Grout Condition. That is,

\[
\theta_i = g_{i,1} \gamma_1 + \cdots + g_{i,q} \gamma_q.
\]

The parameter \( g_{i,j} \) is the probability that a component with grout condition \( i \) will result in test result \( j \). There are many paths through the network that can result in such a match between condition and outcome. For example, the variable pitting has 4 possible states; surface corrosion and internal corrosion both have two states; condition has 3 states; installation has 2 states; and capacity has 4 states so for a particular state of grout there are 384 different combinations of these states that will correspond to any test outcome. As such, during the elicitation exercise we must elicit the conditional probability tables for each node to assess this link.

In a full Bayesian analysis, having defined all of the relevant probability distributions, we could utilise Bayes Theorem to minimise the expected posterior variance of the probabilities of being in each Grout Condition state. Instead, consider the associated Bayes linear Bayes directed graphical model. In order to fully quantify this model we need the prior expectation and variance of \( \theta, E_0(\theta), \text{Var}_0(\theta) \), the prior expectation and variance of \( Y, E_0(Y), \text{Var}_0(Y) \) and the prior covariance between them \( \text{Cov}_0(\theta, Y) \).

We can express the prior moments of \( \theta \) in terms of those of \( Y \) using the relationship above. Thus, if we can evaluate all of the moments of \( Y \), this gives us all of our prior specifications. The following theorem indicates how this can be achieved.
Theorem 1  The prior mean and variance of $\gamma_i$ are given by
\[
E_0(\gamma_i) = \frac{\alpha_i}{\alpha_i + \beta_i} \prod_{j=1}^{i-1} \frac{\beta_j}{\alpha_j + \beta_j}
\]
\[
\text{Var}_0(\gamma_i) = \frac{\alpha_i(a_i+1)}{(\alpha_i + \beta_i)(a_i + \beta_i + 1)} \prod_{j=1}^{i-1} \frac{\beta_j(\beta_j + 1)}{(\alpha_j + \beta_j)(\alpha_j + \beta_j + 1)} - \left(\frac{3a_i}{\alpha_i + \beta_i}\right)^2 \prod_{j=1}^{i-1} \frac{\beta_j}{\alpha_j + \beta_j}
\]
for all $i = 1, ..., q$. The prior covariance between any $\gamma_i, \gamma_j$, for $i < j$, is given by
\[
\text{Cov}_0(\gamma_i, \gamma_j) = \frac{\alpha_i \beta_i}{(\alpha_i + \beta_i)(a_i + \beta_i + 1)} \frac{\alpha_j}{(\alpha_j + \beta_j)(a_j + \beta_j + 1)} \prod_{k=1}^{j-1} \frac{\beta_k(\beta_k + 1)}{(\alpha_k + \beta_k)(\alpha_k + \beta_k + 1)} \times \prod_{i=i+1}^{j-1} \frac{\beta_i}{(\alpha_i + \beta_i)} - \frac{\alpha_i}{\alpha_i + \beta_i} \frac{\alpha_j}{a_j + \beta_j} \prod_{m=1}^{i-1} \frac{\beta_m}{\alpha_m + \beta_m} \prod_{n=1}^{j-1} \frac{\beta_n}{\alpha_n + \beta_n}
\]
Proof: The proof is given in the Appendix.

Thus we see that all of the prior specifications in the model, including that of the prior covariance matrix, are made simply as a result of identifying the parameter values of the marginal beta distributions $(\alpha_i, \beta_i)$ from the experts’ elicited beta distributions for the conditional probabilities.

Now that we have made all of the necessary prior specifications we can use Bayes Theorem to make conjugate updates of $\theta$ upon observation of $X = x$. This results in the posterior expectation and variance of $E(\theta | x), \text{Var}(\theta | x)$, which are found using the parameter values in (3) and the equations in Theorem 1.

We can use Bayes linear kinematics to update the expectations and variances of the probabilities of the different Grout Condition states. This gives:
\[
E(\gamma; \theta) = E_0(\gamma) + \text{Cov}_0(\gamma, \theta)\text{Var}_0^{-1}(\theta)[E(\theta | x) - E_0(\theta)]
\]
\[
\text{Var}(\gamma; \theta) = \text{Var}_0(\gamma) + W_0(\gamma, \theta)\text{Var}(\theta | x)W_0^T(\gamma, \theta)
\]

We then evaluate this adjusted variance over all of the possible data realisations and evaluate the minimum as the optimal test strategy as in (6).

3.3 Results

Initially we consider the probability specifications of a single expert, who we shall call Expert 1, for Grout Condition, which has three categories; fully effective, partially effective and not effective. The prior specifications made by Expert 1 for Grout Condition are of her median, $q_{0.5}$, and “surprise limits”, which are judged to be 5% and 95% quantiles, $q_{0.05}, q_{0.95}$, for the probability that Grout Condition is fully effective and partially effective given that it is not fully effective. The probability that Grout Condition is not effective given that it is not fully effective or partially effective must be 1.
The prior specifications made by Expert 1 and the beta distribution parameter values which approximately satisfy these conditions are given in Table 1. The beta distribution values were found using software package SHELF (http://www.tonyohagan.co.uk/shelf/).

Table 1: The elicited quantiles from Expert 1 and the resulting beta distribution parameter values.

<table>
<thead>
<tr>
<th>Condition</th>
<th>(q_{0.05})</th>
<th>(q_{0.5})</th>
<th>(q_{0.95})</th>
<th>(\alpha_i)</th>
<th>(\beta_i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fully Effective</td>
<td>0.45</td>
<td>0.5</td>
<td>0.8</td>
<td>1.34</td>
<td>1.09</td>
</tr>
<tr>
<td>Partially Effective</td>
<td>0.8</td>
<td>0.95</td>
<td>1</td>
<td>1.95</td>
<td>0.315</td>
</tr>
<tr>
<td>Not Effective</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

This leads to prior unconditional expectations and variances of being in each of the three categories, for Expert 1, as in Table 2.

Table 2: The resulting prior expectations and variances from Expert 1 for the unconditional probabilities of being in each category.

<table>
<thead>
<tr>
<th>Condition</th>
<th>(\mathbb{E}_0(\gamma_i))</th>
<th>(\text{Var}_0(\gamma_i))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fully Effective</td>
<td>0.407</td>
<td>0.0563</td>
</tr>
<tr>
<td>Partially Effective</td>
<td>0.510</td>
<td>0.0567</td>
</tr>
<tr>
<td>Not Effective</td>
<td>0.0824</td>
<td>0.0160</td>
</tr>
</tbody>
</table>

We see from the table that grout condition is most likely to be partially effective in the prior, followed by fully effective and is least likely to be not effective. The uncertainty on these probabilities follows the same pattern, being largest for partially effective, followed by fully effective and finally not effective.

There are 6 possible tests which could be carried out; we label these tests 1-6. For each we calculate the values of \(g_{1,j}\), the terms representing all of the possible routes through the network between Grout Condition and the test in question. We use the mean value of \(g_{1,j}\) in the analysis. For example, for Test 1, these coefficients are given by \(g_{1,1:1,2,3} = (0.894,0.888,0.865)\) and \(g_{2,1:1,2,3} = (0.106,0.112,0.135)\).

We can then use these values and carry out the steps detailed in Section 3.2 for each of the tests to find the information gain from each test \(t\), represented by the expected reduction in variance from performing the test. For Expert 1, we take minus the logarithm of this quantity for each of the tests and plot them in Figure 3.

The lower the ratio of posterior uncertainty to prior uncertainty, the more information has been gained and the larger the bar will be on the plot. We see that, using the data from Expert 1, Test 5 is the optimal test to perform, followed by Test 6. That is, these two tests have the smallest posterior variance. The other four tests are performing relatively poorly using this criteria and all have fairly similar information gain and relatively large posterior variances. There is a clear step change in information gain between Tests 5 and 6 and the other four tests.
Figure 3: A Bar Chart representing the Information Gain from each of the Tests for Expert 1.

We can perform this analysis for the other three experts from whom the probabilities to quantify the BN were elicited to produce their estimates of uncertainty reduction for each of the 6 tests. These values have been plotted directly (without the log transformation) for each combination of expert and test in Figure 4.

Figure 4: A Plot representing the Information Gain from each of the Tests for all Four Experts.
The four groups of tests represent the four experts. As the log transformation has not been made in this case we are considering the ratio of posterior variance to prior variance directly and so lower values represent better tests. We see that all four experts assess Test 5 and Test 6 to be two of the best tests to perform. There is therefore consensus from the experts for these tests.

There is disagreement between the experts, however, in the effectiveness of Test 3 and Test 4, with Expert 2 assessing them to be as effective as Test 5 and Test 6, Expert 1 assessing them to be the least effective tests and Experts 3 and 4 rating them somewhere between these two extremes. All of the experts are agreed that Test 1 and Test 2 are not effective tests for assessing the Grout Condition.

We can also rank each of the tests for each of the experts using the values we have calculated for the expected information gain for each of the tests. These rankings are given in Table 3.

Table 3: The Rankings of the 6 Tests from each of the Experts based on the Expected Information Gain Criteria.

<table>
<thead>
<tr>
<th>Expert/Test</th>
<th>Test 1</th>
<th>Test 2</th>
<th>Test 3</th>
<th>Test 4</th>
<th>Test 5</th>
<th>Test 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

From the table we can see the divides between the preferences over the tests even more clearly. Test 5 and Test 6 are the most effective tests in the eyes of all four experts, with three identifying Test 6 as optimal and one Test 5. Except for Expert 1, all of the experts then identify Tests 3 and 4 as the next best tests to perform, with all preferring Test 3 to Test 4. Expert 1 also finds Test 3 to be a better test than Test 4.

Experts 2, 3 and 4 all assess Test 1 and Test 2 to be the worst tests for gaining information about the state of Grout Condition, with Experts 2 and 3 identifying Test 1 as the worst test and Expert 4 identifying Test 2 as the worst test.

In actual fact, the company which the example is based on chose to carry out Test 6.

4 Conclusions

We have investigated the choice of test procedure in a physical engineering system under uncertainty. The method proposed is to use a BN. This allows relationships between possible tests and the physical state of the system to be structured qualitatively and quantified in a user-friendly and rigorous manner. The optimal test is the one which provides the maximum information about important measures of the underlying state of the system.

Evaluating the expected information gain from performing a test requires the prior predictive distribution over the relevant variables. Acquiring this is a non-trivial task and so an approximation based on Bayes linear kinematics has been proposed. This has advantages in terms of speed and should be close to the full Bayesian solution when the Bayesian model is fairly linear. The methodology has been illustrated using a case study based on a real industrial engineering system.

In the case study we saw that the information gain derived from the Bayes linear Bayes graphical model could be used, for each expert, to identify that expert’s optimal test, or, perhaps more usefully, the tests which were close to optimal. This allowed the identification of tests which all experts felt were useful.
Future work could go in various directions. There is work on aggregating expert judgements when multiple experts are used. An alternative to providing a separate solution for each expert would be to find a single aggregated measure of information gain which represents all of the experts.

Previous work into Bayes linear Bayes graphical models and Bayes linear kinematics [10, 15] has proposed transforming quantities such as probabilities and rates before performing Bayes linear updates in order to improve approximations. It would be interesting to consider transformations such as taking logits in this case.

More generally, performance of an engineering system may be linked to the function of the system and therefore not constrained to physical components. The use of BNs to find the optimal test based on information gain, or other criteria, can incorporate this extra level of complexity into the decision making process. Future work could examine this.

References


Appendix

Proof. The expectation of $\gamma_i$ can be found as the expectation of $\prod_{j=1}^{q}(1 - \mu_j) \times \mu_j$. This is then

$$E(\gamma_i) = \int_0^1 \cdots \int_0^{\beta_i} \prod_{j=1}^{q}(1 - \mu_j) \times \mu_j \frac{\Gamma(\alpha_k + \beta_k)}{\Gamma(\alpha_k)\Gamma(\beta_k)} \mu_k^{\alpha_k-1}(1 - \mu_k)^{\beta_k-1} d\mu_1 \cdots d\mu_q$$

$$= \frac{\alpha_i}{\alpha_i + \beta_i} \prod_{j=1}^{q} \left\{ \frac{\beta_j}{\alpha_j + \beta_j} \right\}$$

In order to calculate the variance of $\gamma_i$, we first need the second moment. This is

$$E(\gamma_i^2) = \int_0^1 \cdots \int_0^{\beta_i} \prod_{j=1}^{q}(1 - \mu_j)^2 \times \mu_j^2 \prod_{k=1}^{q} \frac{\Gamma(\alpha_k + \beta_k)}{\Gamma(\alpha_k)\Gamma(\beta_k)} \mu_k^{\alpha_k-1}(1 - \mu_k)^{\beta_k-1} d\mu_1 \cdots d\mu_q$$

$$= \frac{\alpha_i(a_i + 1)}{(\alpha_i + \beta_i)(\alpha_i + \beta_i + 1)} \prod_{j=1}^{q} \left\{ \frac{\beta_j(\beta_j + 1)}{(\alpha_j + \beta_j)(\alpha_j + \beta_j + 1)} \right\}$$

The variance result follows immediately from this. In order to calculate the covariance we need the cross moment between $\gamma_i \gamma_j$. This is

$$E(\gamma_i \gamma_j) = \int_0^1 \cdots \int_0^{\beta_i} \prod_{k=1}^{q}(1 - \mu_k)^2 \prod_{l=1}^{\beta_i}(1 - \mu_j) \times \mu_j \prod_{m=1}^{q} \frac{\Gamma(\alpha_m + \beta_m)}{\Gamma(\alpha_m)\Gamma(\beta_m)} \mu_m^{\alpha_m-1}(1 - \mu_m)^{\beta_m-1} d\mu_1 \cdots d\mu_q$$

$$= \prod_{k=1}^{\beta_k} \left\{ \frac{\beta_k}{\alpha_k + \beta_k} \right\} \prod_{l=1}^{\beta_i} \left\{ \frac{\beta_l}{\alpha_l + \beta_l} \right\} \prod_{m=1}^{\beta_j} \left\{ \frac{\beta_m}{\alpha_m + \beta_m} \right\} \times \frac{\alpha_i\beta_j}{(\alpha_i + \beta_i)(\alpha_j + \beta_j + 1)}$$

The covariance result follows immediately from this.

Kevin J. Wilson has an MMathStat and Ph.D. in Bayesian Statistics from Newcastle University. He has worked in the Management Science Department at the University of Strathclyde since May 2011 as a Research Associate in Risk and Reliability. His research interests include the elicitation of prior information, Bayes linear methods and Bayesian inference. He is a fellow of the Royal Statistical Society and member of the Operational Research Society.

John Quigley gained a B.Math in Actuarial Science from the University of Waterloo, Canada and a Ph.D. in Management Science from the University of Strathclyde, where he is now a Professor. John’s main research interests are in elicitation of prior distributions, statistical inference and applied probability modelling with application areas in risk and reliability. He has experience with military, aerospace and railway industries. John is an Associate of the Society of Actuaries, a Chartered Statistician, and a member of the Safety and Reliability Society and the Institute of Mathematical Statistics.

Tim Bedford is a Professor in the Department of Management Science at the University of Strathclyde. He studied mathematics at Warwick University where he obtained BSc, M.Sc. and Ph.D. degrees. He holds an Honorary Doctorate from the Faculté Polytechnique de Mons and is a Fellow of the Royal Society of Edinburgh. He held a fellowship at King’s College Cambridge where he was a member of the King’s College Research Centre project on Chaotic Dynamics. He took a lectureship in applied probability at Delft University of Technology in the Netherlands, and became Senior Lecturer in Applications of Decision
Theory. Current research interests are in probabilistic risk analysis, uncertainty modelling and decision analysis.

Lesley Walls is a Professor in the Department of Management Science at the University of Strathclyde. Lesley is a Chartered Statistician, a Member of the Operational Research Society and an elected Fellow of the Safety and Reliability Society. She is an Associate Editor IEEE Transactions on Reliability and a member of the Editorial Board of the Journal of Risk and Reliability. She was Co-chair Mathematical Methods in Reliability Conference held in Glasgow 2007. She has been involved in consultancies for organisations such as the MoD, DSTL, Railway Safety, BAE SYSTEMS, Goodrich and Siemens. She is a UK Expert on Reliability to International Electrotechnical Committee (IEC TC56 WG2).